Fuzzy-Clustering als methodische Grundlage zur dynamischen Marktsegmentierung

vom Fachbereich IV
– Mathematik, Naturwissenschaften, Wirtschaft und Informatik –
der Universität Hildesheim

zur Erlangung des Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

von
Anneke Minke
aus Hannover

Versicherung gemäß §3 Abs. 2a der Promotionsordnung in der Beschlussfassung vom 11.07.1991

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit mit dem Titel

Fuzzy-Clustering als methodische Grundlage
zur dynamischen Marktsegmentierung

im Bereich Betriebswirtschaft mit Schwerpunkt im Marketing selbstständig und ohne unerlaubte Hilfe verfasst und die benutzten Hilfsmittel vollständig angegeben habe.

Ich habe bisher an keiner in- oder ausländischen Universität einen Antrag auf Zulassung zur Promotion gestellt noch die vorliegende oder eine andere Arbeit als Dissertation vorgelegt.

Hannover, 03. Mai 2013

Anmerkungen

Alle Aussagen in dieser Arbeit wurden sorgfältig geprüft und werden ggf. anhand von Quellenangaben belegt. Tabellen, Abbildungen und Übersichten fremder Autoren wurden als solche gekennzeichnet. Sofern keine Quelle angeführt wurde, handelt es sich um selbst erstellte Inhalte, eine gesonderte Kennzeichnung erfolgt in diesem Fall nicht.

Sämtliche Experimente auf Basis künstlich generierter Daten wurden unter Anwendung eines eigens dafür entwickelten Java-Applets durchgeführt, das auf Anfrage unter mail@annekeminke.de zu bekommen ist.

Kurzfassung

Das Aufdecken zeitabhängiger Änderungen in Datenstrukturen im Bereich der Marktforschung ist von grundlegender Bedeutung, um gezielt für die Zukunft planen zu können. Insbesondere die Veränderung einzelner Marktsegmente muss nachvollzogen werden, damit ein Unternehmen auf Entwicklungen reagieren und entsprechende Maßnahmen einleiten kann. Eine ausschließlich deskriptive Untersuchung der aufgetretenen Strukturänderungen ist jedoch nicht ausreichend; vielmehr müssen Trends in die Planung einbezogen werden, da nur so rechtzeitig eine Anpassung der Marketinginstrumente vorgenommen werden kann, um ggf. Entwicklungen zu fördern oder ihnen – sofern möglich und sinnvoll – entgegenzuwirken. Das Change Mining bietet Ansätze, Strukturen im dynamischen Kontext zu analysieren und so zeitabhängige Veränderungen frühzeitig aufzudecken.

Zum automatischen Erkennen von Marktsegmenten ist die Clusteranalyse ein weit verbreitetes Verfahren; sie ermöglicht die Einteilung eines heterogenen Datensatzes, z.B. verschiedener Kundenverhaltensdaten, in homogene Teilgruppen. Bei der Analyse dynamischer Änderungen innerhalb der Daten ist außerdem das Aufdecken gradueller Unterschiede zwischen den einzelnen Untersuchungszeitpunkten von hoher Bedeutung, um zukünftige Entwicklungen und Trends prognostizieren zu können. Daher erscheint eine harte Clustereinteilung in diesem Kontext ungeeignet, vielmehr werden Zugehörigkeitsgrade benötigt. Aus diesem Grund empfiehlt sich die Fuzzy-Clusteranalyse als Grundlage für die dynamische Betrachtung von Marktsegmenten.

In der vorliegenden Arbeit wird aufbauend auf einer allgemeinen Einführung in die Thematik des Data Minings und des Change Minings eine kurze Vorstellung von Ansätzen zur Untersuchung dynamischer Veränderungen unter Anwendung anderer, in der Marktforschung relevanter Data Mining-Verfahren vorgenommen, bevor eine Einführung in die Fuzzy-Clusteranalyse im Speziellen erfolgt. Aufgrund des Untersuchungsziels, dynamische Veränderungen innerhalb einer Clusterstruktur aufzudecken, werden neue Ansätze zur possibilistischen Fuzzy-Clusteranalyse erläutert, die auf Homogenität innerhalb einzelner Cluster ausgerichtet sind und unabhängig von möglichen zeitabhängigen Änderungen angewendet werden können.

Zum Nachverfolgen aufgetretener Veränderungen und der Vorhersage möglicher Entwicklungen werden verschiedene Maße vorgestellt und genauer betrachtet, um den Prozess des Change Minings zu verdeutlichen. Dabei wird zwischen der Analyse gradueller Änderungen innerhalb einzelner Cluster und abrupter Veränderungen bzgl. der Gesamtstruktur und der darin vorhandenen Clusterzahl unterschieden. Die Ergebnisse der Untersuchung gradueller Änderungen bzgl. Größe, Dichte und Volumen der individuellen Cluster werden dabei zur Ermittlung abrupter Änderungen bzgl. der Clusterzahl, d.h. neu entstandener, zu elimierender, vereinigter oder geteilter Cluster hinzugezogen, da sie diese implizieren können. Das Vorgehen zur Analyse einzelner Veränderungsarten wird jeweils anhand detailliert dargestellter Experimente veranschaulicht, die mit Hilfe eines eigens für diesen Zweck entwickelten Java-Applets durchgeführt wurden.

Abstract

In market research problems, detecting time-dependent changes in data structures is of essential importance for systematic future planning. Especially understanding changes concerning separate market segments is a fundamental task for cooperations in order to react to specific developments and to initiate appropriate actions. However, an exclusively descriptive analysis of existing structural changes is not sufficient, but instead emerging trends have to be included in planning strategies to modify different marketing tools for promoting certain developments or – if possible and reasonable – to counteract them. The field of Change Mining is concerned with approaches for analyzing structures in a dynamic environment and hence, to detect time-dependent changes at an early stage.

Cluster analysis is a widely used technique for identifying market segments automatically as it enables dividing a heterogeneous data set, e.g. data regarding customer behavior, into homogeneous subsets. Additionally, when analyzing time-dependent changes within a data set, the detection of gradual differences between different points in time is essential for predicting future developments and trends. Therefore, hard clustering techniques are of limited use as membership degrees are needed rather than crisp cluster assignments. Hence, fuzzy clustering appears suitable as basis for the dynamic examination of market segments.

In this work, a general introduction to Data Mining and Change Mining is given and on this basis, Data Mining approaches for analyzing time-dependent changes are presented which are relevant for different market research tasks. Afterwards, the fundamentals of fuzzy clustering are introduced. Due to the focus on detecting changes in a general cluster structure, new approaches for possibilistic fuzzy clustering are defined which emphasize the homogeneity of individual clusters and which are independent of possible changes that may occur over time.

For tracking emerging changes and predicting potential developments, different measures are introduced and evaluated for demonstrating the Change Mining process. Thereby, gradual changes within individual clusters and abrupt changes concerning the overall cluster structure are considered separately. The results concerning gradual changes with regard to size, density, and volume of particular clusters are consulted when investigating abrupt changes relative to the number of clusters, i.e. newly developed, eliminated, merged, or split clusters, as they can indicate abrupt changes. The analysis of these types of changes is demonstrated using detailed illustrations of experimental results. The experiments were conducted by employing a Java applet which was exclusively developed for this purpose.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Algorithmenverzeichnis

We are drowning in information and starving for knowledge.
R. D. Roger

1 Einführung

In einer vor einigen Jahren noch nicht vorstellbaren Menge werden in der heutigen Zeit Daten gespeichert, die ihrerseits den verschiedensten Quellen entstammen. Beispielhaft anführen lassen sich hierzu Scannerkassen in Supermärkten, Online-Shops mit den dazugehörigen Transaktionen sowie Weblogs, die das Verhalten eines Nutzers im Internet aufzeichnen. Dieses Vorgehen führt zu Datenmengen in enormer Größe; einige Beispiele sollen diesen Umstand greifbar machen:

- e-Commerce-Unternehmen können wöchentlich bis zu mehreren hundert Megabyte an Daten erfassen, anhand derer das Nutzungsverhalten der Kunden nachvollzogen werden kann (vgl. Wiedmann u. a., 2003, S. 21).
- Der US-amerikanische Telekommunikationskonzern AT&T verfügt zur Zeit über 105,2 Mio. Kunden. Die zugehörigen Informationen, auch bzgl. des Telefonierverhaltens etc., werden zwecks Weiterverarbeitung in Multiterabyte-Datenbanken abgelegt (vgl. AT&T, 2013; Cios u. a., 2007, S. 4).

Die Speicherung der erzeugten Datenmassen erfolgt häufig in sogenannten Data Warehouses. Viele Unternehmen verfügen über große, unter anderem mit Kundeninformationen gefüllte Data Warehouses; dabei kann ein verhältnismäßig kleines Data Warehouse mehr als 100 Millionen Transaktionen enthalten (vgl. Bramer, 2007, S. 1). Einen Nutzen erhalten diese Daten jedoch nur dann, wenn sie auch analysiert und ausgewertet werden. Mit den Fortschritten der Technik, die das Sammeln und Speichern der Daten zu geringen Kosten fördert, wuchs auch die Erkenntnis, dass in diesen Datenmengen Wissen lagert, das sich zu extrahieren lohnt (vgl. z.B. Decker, 1993, S. 45ff.). So können Unternehmen aus den gesammelten Transaktionsdaten beispielsweise interessante Zusammenhänge zwischen den Kunden und den von ihnen gekauften Produkten aufdecken. Dabei wird das Ziel verfolgt, das eigene Unternehmen im Wettbewerb zu stärken und damit profitabler zu machen. Eine manuelle Analyse erweist sich jedoch als zeitaufwändig, teuer und subjektiv, zudem bei der vorhandenen Datenmenge – Datenbanken mit Daten zu $n = 10^9$ Objekten sind heute keine Seltenheit mehr – als nicht praktikabel (vgl. Fayyad u. a., 1996; Freitas, 2002). Um diesem Umstand zu begegnen, wurden in der Vergangenheit Algorithmen entwickelt, die in den verfügbaren Daten nach Mustern suchen und vorgegebene Strukturen prüfen. Der Prozess der Datenanalyse wird allgemein als Knowledge Discovery in Databases (KDD), das Anwenden eines spezifischen Algorithmus auf die Daten als Data Mining bezeichnet.

1.1. Knowledge Discovery in Databases

Der Begriff *Knowledge Discovery in Databases* (KDD) wurde erstmals auf dem ersten KDD-Workshop 1989 verwendet (vgl. Fayyad u. a., 1996). Durch diesen Begriff sollte die Fokussierung auf das Wissen als Endprodukt einer datengesteuerten Analyse gelegt werden. Bis heute ist dieser Begriff insbesondere im Bereich der Künstlichen Intelligenz sowie des Maschinellen Lernens weit verbreitet. Der KDD-Komplex entstand im Überschneidungsbereich verschiedener Forschungsgebiete, neben den bereits genannten Gebieten der Künstlichen Intelligenz und des Maschinellen Lernens spielten hier vor allem Statistik, Pattern Recognition im Allgemeinen sowie die Datenbankentwicklung eine entscheidende Rolle. Auch heute noch beschreibt Knowledge Discovery in Databases ein interdisziplinäres Feld, das Methoden verschiedener Gebiete verwendet (vgl. Freitas, 2002). Abstrahiert liegt das Ziel der KDD in der Entwicklung und Anwendung von Methoden, die darauf ausgerichtet sind, in unübersichtlichen Daten verständliche Muster aufzudecken. Dabei werden Theorien und vor allem Werkzeuge benötigt, die den Nutzer dabei unterstützen, nützliche Informationen und damit *Wissen* aus schnell wachsenden Datenmengen zu extrahieren (vgl. Fayyad u. a., 1996). Bramer (2007, S. 2) definiert die Wissensentdeckung als

> „non-trivial extraction of implicit, previously unknown and potentially useful information from data."

In diesem Zusammenhang zeigt der Begriff *nichttrivial* an, dass es sich dabei nicht um einfache statistische Größen wie die Berechnung eines Mittelwertes o.ä. handelt, sondern dass die Ermittlung komplexer Zusammenhänge gefordert wird. Des Weiteren bedeutet dies, dass neuartige, für die Analyse nützliche Informationen extrahiert werden sollen. Fayyad u. a. (1996) erweitern diese Definition zusätzlich:

> „KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns from data."

Neben den bereits in der erstgenannten Definition geforderten Eigenschaften der Neuartigkeit und der Nützlichkeit sollen die aufgedeckten Muster, d.h. die Strukturen in den Daten, auch über die untersuchten Daten hinaus mit einer gewissen Sicherheit gültig sein; ferner müssen sie verständlich und nachvollziehbar sein. Diese Forderung liegt darin begründet, dass ein Nutzer, der basierend auf einem erlernten Muster Entscheidungen trifft, diese Entscheidungen verstehen muss, um ihnen vertrauen zu können. Zusammenfassend bedeutet dies, dass Ergebnisse anvisiert werden, die eine gewisse vorhersagende Genauigkeit aufweisen und für den Nutzer verständlich sind, damit sie ihm zustatten kommen (vgl. Freitas, 2002).

Der Einsatz verschiedener Verfahren zur Extraktion von Wissen aus großen Datenbeständen, das sogenannte *Data Mining*, stellt lediglich einen einzelnen Schritt des KDD-Prozesses dar (vgl. Abbildung 1.1), der vier wesentliche Schritte umfasst:

1. *Datenintegration*: Aus verschiedenen Quellen, beispielsweise aus verschiedenen Unternehmensbereichen, werden Daten in einem gemeinsamen Datenspeicher (Data Store) zusammengeführt. Hierbei ist darauf zu achten, dass Inkonsistenzen eliminiert werden sowie ein einheitliches Format verwendet wird (vgl. Freitas, 2002).

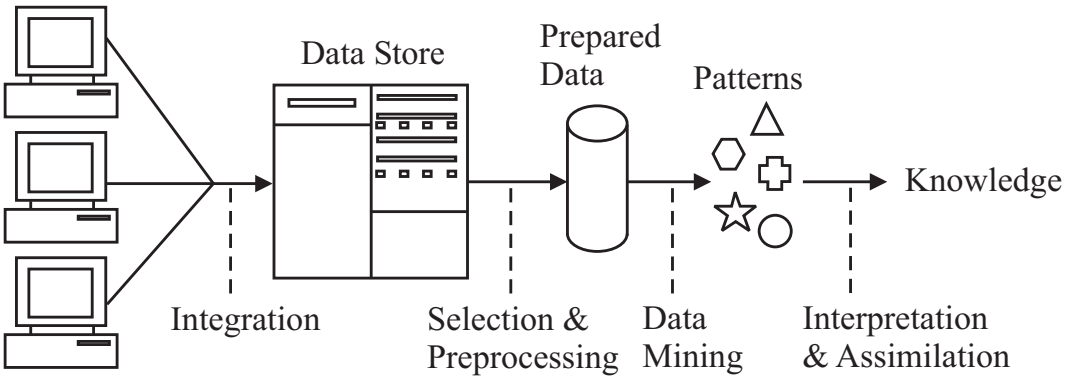


Abbildung 1.1.: KDD-Prozess (Bramer, 2007, S. 2)

2. *Datenselektion und -vorbereitung* (Selection & Preprocessing): Neben einer allgemeinen Datenbereinigung und dem Eliminieren von Rauschen[1] müssen Missing Values beachtet und gegebenenfalls ergänzt werden. Ferner erfolgt eine Auswahl der für den Untersuchungsgegenstand relevanten Eigenschaften und Attribute, da nicht alle vorhandenen Attribute für eine Data Mining-Aufgabe von Bedeutung sind. Irrelevante Attribute können nicht nur den Anwender verwirren, sie können sogar dazu führen, dass für die Erklärung eines Sachverhaltes letztendlich bedeutungslose Zusammenhänge erstellt werden und das resultierende Modell schließlich Fehler aufweist. Bei großen Datenmengen sollte außerdem eine Redzuierung auf repräsentative Teilmengen erfolgen, um den Prozess nicht unnötig zu verlangsamen (vgl. Raghavan und Hafez, 2000).
3. *Data Mining*: Durch die Anwendung spezifischer Algorithmen auf die vorbereiteten Daten (Prepared Data) werden Muster (Patterns) erlernt, die die Strukturen in den Daten widerspiegeln. Hierbei handelt es sich um den Kern des KDD-Prozesses, der im Folgenden näher betrachtet wird.
4. *Interpretation und Anwendung* (Interpretation & Assimilation): Erst durch die Ergebnisinterpretation können die erlernten Muster auch für einen Laien verständlich dargestellt werden. Ferner sollte in diesem Schritt ein Postprocessing stattfinden, um evaluieren zu können, inwiefern die ermittelten Modelle allgemeine Gültigkeit besitzen. Erst im Anschluss erfolgt eine Anwendung der Ergebnisse.

Beim Kernschritt des KDD-Prozesses, dem eigentlichen Data Mining, handelt es sich um einen Sammelbegriff, der verschiedene Techniken zur Mustererkennung und Modellkonstruktion beschreibt (vgl. Chen u. a., 2005). Raghavan und Hafez (2000) bezeichnen diese Phase des KDD-Prozesses als

> „process of discovering potentially valuable patterns, associations, trends, sequences and dependencies in data."

Dieser Prozessschritt verfolgt das Ziel, Zusammenhänge in den vorhandenen Daten aufzudecken. Dieses Vorgehen ermöglicht somit z.B. durch die Analyse der Kunden- und Transaktionsdaten,

[1] Als *Rauschen* in Daten bezeichnet man fehlerhafte Werte, die zufällig, z.B. aufgrund von Messfehlern, entstehen, sowie Werte, die aufgrund der Varianz innerhalb eines Attributs vorhanden sind (vgl. Cios u. a., 2007, S. 42). Bei starkem Rauschen kann dies ohne Bearbeitung den Informationsgehalt des Ergebnisses des KDD-Prozesses negativ beeinflussen.

bestimmte Muster im Kundenverhalten zu identifizieren; als Folge davon können Kunden gezielter angesprochen und damit der eigene Umsatz gesteigert werden. Es besteht weiterhin die Möglichkeit, starke Abweichungen im Kundenverhalten frühzeitig zu erkennen und die Ursachen dafür zu ermitteln; dies ist vor allem im Bereich des Missbrauchs von Kundendaten relevant, um Betrügereien schnellstmöglich aufzudecken (vgl. Fayyad u. a., 1996).

1.2. Verfahren des Data Minings

Im Rahmen des Data Minings werden zwei Verfahrensarten unterschieden: *strukturentdeckende* und *strukturprüfende* Verfahren. Die strukturentdeckenden Verfahren sollen Zusammenhänge zwischen vorhandenen Variablen oder Datensätzen aufdecken, daher werden diese Verfahren auch unter dem Begriff der Interdependenzanalyse zusammengefasst. Welcher Art die gesuchten Zusammenhänge sind, ist von der konkreten Problemstellung und damit dem gewählten Verfahren abhängig. Zu den strukturentdeckenden Verfahren zählen unter anderem die folgenden:

- *Clusteranalyse*
  Ziel der Clusteranalyse ist es, Objekte mit ähnlichen Datensätzen innerhalb der Daten aufzudecken und diese in sogenannte Cluster zu gruppieren. Dies erfolgt auf Basis der Eigenschaftsausprägung einzelner Objekte (vgl. z.B. Backhaus u. a., 2006, S. 619ff.; Cios u. a., 2007, S. 255ff.). Ein Anwendungsbeispiel für die Clusteranalyse stellt die Einteilung von Kunden in einzelne Segmente dar, bei der Kunden z.B. basierend auf ähnlichen Eigenschaftsausprägungen bzgl. ihres Kaufverhaltens zusammengefasst werden, so dass eine gezielte Bearbeitung einzelner Segmente ermöglicht wird.
- *Faktoranalyse*
  Bei der Faktorenanalyse sollen Zusammenhänge zwischen einzelnen, i.d.R. quantitativen Variablen untersucht werden. Hintergrund ist die Überlegung, dass sich aus vielen, ein Objekt beschreibenden Eigenschaften wenige latent vorhandene Faktoren ableiten lassen, die voneinander unabhängig sind (vgl. z.B. Backhaus u. a., 2006, S. 259ff.; Berekoven u. a., 2006, S. 217ff.). Auf diese Weise ist es möglich, die hintergründigen Eigenschaften aufzudecken, die beispielsweise für Konsumenten eines Produktes relevant sind.
- *Assoziationsregeln*
  Die Suche nach Assoziationsregeln befasst sich mit der Analyse von Kundenverhaltensmustern, auf deren Basis gezielt Produkte angeboten und platziert werden können; es gilt also, Abhängigkeiten zwischen Attributwerten in der Form *Kunden, die A kaufen, kaufen auch B* aufzudecken (vgl. z.B. Bramer, 2007, S. 187ff.; Cios u. a., 2007, S.289ff.). Ein Beispiel für die Anwendung der Assoziationsregeln sind Recommender, die Kunden anhand der erlernten Verhaltensweisen weitere Produkte empfehlen.
- *Multidimensionale Skalierung*
  Die Multidimensionale Skalierung (MDS), auch Mehrdimensionale Skalierung genannt, verfolgt das Ziel, eine Repräsentation von (Un-)Ähnlichkeitsmessungen zwischen Objektpaaren als Distanzen im niedrigdimensionierten Raum zu erstellen und so eine visuelle Untersuchung der Objektstruktur zu ermöglichen (vgl. z.B. Berekoven u. a., 2006, S. 223ff.; Borg und Groenen, 2005, S. 3ff.). Auf diese Weise können Wahrnehmungsräume von Kunden visualisiert und auf relevante Eigenschaften für die Beurteilungen einzelner Objekte geschlossen werden.

Aufgrund des Mangels einer als *richtig* anzusehenden Struktur werden die Verfahren der Interdepenzanalyse auch unüberwachte Verfahren genannt. Eine genaue Fehlerbestimmung ist im Gegensatz zu den strukturprüfenden Verfahren i.d.R. nicht möglich; vielmehr werden gesonderte Gütemaße benötigt, anhand derer eine Evaluierung der Ergebnisse erfolgen kann.

Die strukturprüfenden Verfahren dienen der Überprüfung von Hypothesen und vorgegebenen Zusammenhängen, d.h., es handelt sich um Verfahren zur Abhängigkeits- bzw. Dependenzanalyse. Die einzelnen Variablen werden in abhängige und unabhängige Variablen unterteilt, so dass die Wirkung der unabhängigen auf die abhängigen Variablen untersucht werden kann. Folgende Verfahren gehören unter anderem in diesen Data Mining-Bereich:

- *Varianzanalyse*
  Die Varianzanalyse gibt Aufschluss über die Abhängigkeit eines quantitativen Merkmals von nominalen Merkmalen (vgl. z.B. Backhaus u. a., 2006, S. 119 ff.; Berekoven u. a., 2006, S. 214f.). So kann beispielsweise die Verpackung eines Produktes Einfluss auf die Absatzmenge haben.

- *Regression*
  Die Regression liefert eine Funktion, die die Abhängigkeit einer i.d.R. quantitativen Variablen von einer oder mehreren unabhängigen quantitativen Variablen beschreibt (vgl. z.B. Backhaus u. a., 2006, S. 45ff.; Berekoven u. a., 2006, S. 210ff.). Ein Beispiel ist die Beeinflussung des Absatzes eines Produkts durch verschiedene Merkmale wie Preis und Werbeaufwand.

- *Klassifikation*[2]
  Bei der Klassifikation soll eine apriori bekannte Klassenzerlegung von Objekten erklärt werden. Die abhängige nominale Variable beschreibt die Klassenzugehörigkeit, die durch die unabhängigen Variablen begründet wird. Die Analyse der Zerlegung kann z.B. anhand einzelner Merkmale mit Hilfe eines sogenannten Entscheidungsbaumes (vgl. Bramer, 2007, S. 41ff.; Cios u. a., 2007, S. 381ff.) oder basierend auf einer Trennfunktion im Rahmen einer Diskrimanzanalyse erfolgen (vgl. u.a. Backhaus u. a., 2006, S. 155ff.; Berekoven u. a., 2006, S. 216f.). Beispielhaft für die Klassifikation ist die Unterscheidung von Produkten in *Renner* und *Flops* zu nennen. Anhand einer Diskriminanzanalyse kann ermittelt werden, in welcher Form einzelne Produkteigenschaften diese Klasseneinteilung beeinflussen.

Da bei den strukturprüfenden Verfahren die abhängigen Zielvariablen im Vorfeld bekannt sind, kann das Ergebnis einer solchen Analyse direkt evaluiert werden, indem die tatsächlichen Werte mit den durch das jeweilige spezifizierte Modell approximierten verglichen werden. Aufgrund dieser Eigenschaft werden diese Verfahren auch als überwachte Verfahren bezeichnet.

[2]Im ökonomischen Kontext und der Marktforschung wird der Begriff *Klassifikation* häufig synonym mit dem Begriff *Clusteranalyse* verwendet. Dieser Umstand liegt darin begründet, dass durch die Clusteranalyse verschiedene Gruppen in den Daten aufgedeckt werden, die auch als Segmente oder Klassen bezeichnet werden. Diese Art der Klassifikation unterscheidet sich jedoch von der im Zusammenhang der strukturprüfenden Verfahren vorgestellten Klassifikation zur Erklärung einer apriori bekannten Zerlegung der Daten.

1.3. Grundlagen des Change Minings

Bis Ende des zwanzigsten Jahrhunderts wurden Data Mining-Verfahren bis auf wenige Ausnahmen hauptsächlich auf statische Datensätze angewendet; der Untersuchung von Veränderungen in bereits erlernten Mustern kam relativ wenig Beachtung zu (vgl. Chen u. a., 2005; Song u. a., 2001). Dementsprechend erfolgte meist eine wiederholte Anwendung statischer Methoden, um aktualisierte Ergebnisse zu erhalten (vgl. Böttcher u. a., 2009; Song u. a., 2001). Die resultierenden Unterschiede zu den zuvor erlernten Mustern wurden dabei nicht oder nur oberflächlich analysiert. Gerade in der heutigen Zeit, in der aufgrund des technischen Fortschritts die Möglichkeit besteht, große Datenmengen wie z.B. Transaktionsdaten in Online-Shops zu speichern, bietet es sich jedoch an, Veränderungen in den Mustern zu analysieren; so können aus den Ergebnissen konkrete Handlungsmöglichkeiten abgeleitet werden. Data Mining-Ansätze, die sich dieser dynamischen Entwicklung von Mustern widmen, werden unter dem Begriff *Change Mining* zusammengefasst.

Bei der Analyse von Datensätzen ist es von großer Bedeutung, zwischen einzelnen Perioden zu trennen. Ohne eine solche Trennung erhielte man andernfalls, d.h. bei einer gemeinsamen statischen Analyse der über mehrere Perioden hinweg erfassten Daten, ungenaue Ergebnisse; besonders zur Vorhersage zukünftiger Entwicklungen wären die Ergebnisse ungeeignet, da keine Unterscheidung zwischen aktuellen und veralteten Daten stattfindet (vgl. Chakrabarti u. a., 1998; Crespo und Weber, 2005). Veraltete Daten enthalten keine oder nur geringe Informationen bzgl. der aktuellen Situation; so können sich beispielsweise die Einstellungen von Kunden geändert haben. Der Gebrauch des ursprünglich spezifizierten Modells kann in diesem Fall zu nicht akzeptierbaren Entscheidungen führen. Als Beispiel hierfür können die Eigenschaften eines erstrebenswerten Autos dienen (vgl. Rissland und Friedman, 1995): Die vor zwanzig Jahren als optimal geltenden Eigenschaften sind aufgrund von technischen und gesellschaftlichen Entwicklungen längst überholt. Die Durchführung einer auf allen jemals erhobenen Eigenschaften basierenden Clusteranalyse erweist sich nicht nur aufgrund der hohen Datenmenge als enorm aufwändig, sondern führt zudem zu schlechter Clusterqualität, da eine wesentliche Veränderung der Daten stattgefunden hat und die veralteten Daten zu stark gewichtet werden. Die Veränderungen der Charakteristika im Verlauf der Zeit dürfen nicht vernachlässigt werden (vgl. Aggarwal u. a., 2003), sondern sind aufzudecken. Um diesen Ansprüchen zu genügen, ist die verfügbare Datenbasis zu pflegen und ständig zu aktualisieren, damit der Lebenszyklus der betrachteten Objekte, d.h. der Produkte, Kunden o.ä., abgebildet und ihre permanente Entwicklung nachvollzogen werden kann (vgl. Pechoucek u. a., 1999; Raghavan und Hafez, 2000). In dieser sich ändernden Datenbasis können über die Zeit hinweg neue Muster entstehen, die zu einer nichttrivialen Änderung der Gesamtstruktur führen (vgl. z.B. Chakrabarti u. a., 1998; Rissland und Friedman, 1995): Es besteht z.B. die Möglichkeit, dass in einer untersuchten Kundenbasis neue Kundensegmente entstehen, die gezielt bearbeitet werden können. Solche Veränderungen müssen aufgedeckt werden, um die Ergebnisse aktuell und für den Anwender wertvoll zu erhalten (vgl. Chakrabarti u. a., 1998). Ziel des Change Minings ist dabei das Erfassen von Trends und zuvor unbekannten, neu entstandenen Teilmustern in der Datenstruktur auf möglichst effiziente Weise (vgl. u.a. Agrawal und Psaila, 1995; Dong u. a., 2003), so dass zukünftige Entwicklungen prognostiziert werden können. Unter anderem können so die Ergebnisse einer entsprechenden dynamischen Analyse herangezogen werden, um die Zahlungsfähigkeit bzw. den Verzug von Kunden im Kreditgeschäft besser vorhersagen zu können (vgl. Sanz Sáiz, 2005). Die

Veränderung muss dabei explizit beschrieben werden, denn nur so kann analysiert werden, welche Veränderungen relevant sind und welche eher Zufallscharakter aufweisen (vgl. Wang u. a., 2003).

Böttcher u. a. (2008) führen den Begriff des Change Minings folgendermaßen ein:

> „Change Mining [is] data mining over a volatile, evolving world with the objective of understanding change."

Die Herausforderung besteht somit darin, auftretende Veränderungen zu erfassen und zu verstehen, um dann die Art der Änderung und vor allem den Zeitpunkt diagnostizieren zur können. Die Autoren gehen in ihrer Definition noch weiter:

> „Change mining is a data paradigm for the study of time-associated data. Its objective is the discovery, modelling[,] monitoring[,] and interpretation of changes in the models that describe an evolving population."

Beim Change Mining handelt es sich dementsprechend um die vollständige Analyse zeitabhängiger Daten zur Abbildung ihrer Entwicklung. Dabei ergibt sich die Schwierigkeit, eine solche Analyse sinnvoll zu gestalten. Die einfachste Möglichkeit besteht darin, regelmäßig eine unabhängige, neue Analyse auf Basis aktueller Daten durchzuführen. Dieses Vorgehen erweist sich jedoch als wenig effizient, vielmehr sollte bereits gelerntes Wissen in die Analyse einbezogen werden (vgl. Raghavan und Hafez, 2000). Auf diese Weise kann z.B. eine bereits gelernte Basismenge an Assoziationsregeln einbezogen werden: Sie gibt Aufschluss über die in der Vergangenheit verstärkt aufgetretenen Zusammenhänge. Die Veränderung der einzelnen Regeln kann zum Aufdecken von Trends beitragen. Bereits erlerntes Wissen kann ferner dazu verhelfen, den allgemeinen Aufwand zu reduzieren, und zu einer Verbesserung der Ergebnisse führen. So kann z.B. der Aufwand einer Clusteranalyse verringert werden, indem vorab anhand der neuen Daten potentielle Änderungen ermittelt werden, bevor die Analyse mit einer geeigneten Clusterzahl und entsprechenden Ergebnissen der Vorperioden initialisiert wird. Zusammengefasset ergeben sich damit drei Möglichkeiten, mit Veränderungen umzugehen (vgl. Crespo und Weber, 2005):

1. Vernachlässigen, d.h. Weiterverwendung des Initialsystems ohne Update:
   Hierbei handelt es sich vom Aufwand her um das günstigste Vorgehen, da keine weiteren Analysen benötigt werden. Dieses Vorgehen birgt jedoch die Gefahr, aufgrund veralteter Muster unzureichende Ergebnisse zu erhalten; ferner können keinerlei Tendenzen aufgedeckt werden.

2. Entwicklung eines neuen Systems nach einer festen Anzahl an Perioden mit Hilfe bekannter statischer Methoden:
   Die hier erfolgende Bestimmung aktueller Strukturen erfordert einen hohen Rechenaufwand, da zu jedem Untersuchungszeitpunkt eine komplett neue Analyse benötigt wird, losgelöst von vorherigen Ergebnissen. Außerdem lässt dieses Vorgehen keine Rückschlüsse auf Entwicklungen zu – sowohl vergangener als auch zukünftiger Art.

3. Update des Initialsystems auf Basis neuer Daten:
   Dieses Vorgehen erweist sich vom Rechenaufwand her günstiger als die zweite Alternative. Weiterhin besteht die Möglichkeit, tatsächliche Veränderungen nachzuvollziehen.

Bevor ein Datensatz im dynamischen Kontext analysiert werden kann, muss eine konkrete Problemdefinition erfolgen. Dabei ist insbesondere die Frage zu beantworten, ob das Resultat beschreibenden oder vorhersagenden Charakter aufweisen soll (vgl. Böttcher u. a., 2008). Sollen ausschließlich die auftretenden Unterschiede zwischen Mustern verschiedener Zeitpunkte herausgestellt und interpretiert werden, handelt es sich um eine Beschreibung: Es wird lediglich das Ergebnis der Veränderung herausgearbeitet. Bei einer Vorhersage handelt es sich hingegen um eine Analyse von höherer Komplexität, da in diesem Fall auch zukünftige Entwicklungen hervorgehoben werden sollen. In diesem Fall soll der Prozess der Veränderungen nachvollzogen und zudem prognostiziert werden, inwiefern sich das nächste Muster von den bisherigen unterscheiden wird. Die Vorhersage ist der Beschreibung übergeordnet, da sie die Beschreibung der aufgetretenen Veränderungen als Bestandteil enthält (vgl. Böttcher u. a., 2008). Weiterhin muss das Hintergrundwissen, das in der Analyse Anwendung finden kann, vorab definiert werden (vgl. Pechoucek u. a., 1999); dies umfasst z.B. die vorher festzulegende Clusterzahl im Rahmen einiger Clusteranalyseverfahren o.ä.

Für eine dynamische Analyse muss vorab festgelegt werden, welche Veränderungen überhaupt auftreten und für den Untersuchungsgegenstand relevant sein können. Liegen zeitbezogene Daten vor, können im Allgemeinen zwei Arten von Veränderungen unterschieden werden: *graduelle Veränderungen* und *abrupte Veränderungen* (vgl. z.B. Rissland und Friedman, 1995; Wang u. a., 2003). Unter graduellen Veränderungen werden diejenigen Veränderungen zusammengefasst, die die charakteristischen Eigenschaften der untersuchten Instanzen betreffen, möglich wären dabei Modifikationen bzgl. der Attribute in Assoziationsregeln oder eine Abweichung in den für ein Segment als typisch angesehenen Eigenschaftsausprägungen. Abrupte Veränderungen betreffen hingegen das gesamte Muster bzgl. der Anzahl darin enthaltener Teilmuster wie neu entstandene oder verschwundene Teilmuster (vgl. Chen u. a., 2005), dazu gehören z.B. das Auftreten neuer Assoziationsregeln oder aber das Verschwinden von Kundensegmenten.

Nach Erreichen der allgemeinen Problemdefinition müssen die für die Analyse benötigten zeitlichen Parameter festgelegt werden, insbesondere die Bestimmung der Zeitintervalle, in denen eine Analyse stattfinden soll, sowie des generell unterliegenden Zeitmodells, das Aufschluss darüber gibt, wie mit älteren Daten zu verfahren ist. Die Länge der Zeitintervalle erweist sich dabei als anwendungsabhängig: So kann es am Aktienmarkt erforderlich sein, mehrmals täglich ein zuvor bestimmtes Muster und das darauf basierende Vorhersagemodell anzupassen, während die Klasseneinteilung von Kreditnehmern und ihre Entwicklung vergleichsweise selten aktualisiert werden muss. Vorab erfolgt somit die Festlegung von Zeitfenstern zur Erfassung der Daten und am Ende einer jeweiligen Periode eine Aktualisierung des zugrunde liegenden Musters (vgl. u.a. Agrawal und Psaila, 1995; Rissland und Friedman, 1995). Außerdem besteht die Möglichkeit, für eine Analyse ältere Daten heranzuziehen, die bereits in eine früher erfolgte Aktualisierung einbezogen wurden, jedoch über einen noch bis in die Gegenwart reichenden Einfluss verfügen. Das Ausmaß dieses Einflusses wird durch das zugrunde liegende Zeitmodell vorgegeben: Es können entweder sich überlappende Zeitfenster verwendet werden, so dass alle Daten, die in dem untersuchten Zeitraum erhoben wurden, unabhängig von ihrem Alter denselben Einfluss haben (vgl. z.B. Angstenberger, 2000, S. 29f.; Crespo und Weber, 2005; Zhou u. a., 2008), oder aber es erfolgt die Anwendung einer exponentiellen Alterungsfunktion, die den einzelnen Instanzen eine alterspezifische Gewichtung zuweist (vgl. Böttcher u. a., 2008).

Abschließend muss der Prozess des Monitorings entworfen werden, um die eigentlichen Struk-

turveränderungen aufdecken und evaluieren zu können (vgl. u.a. Angstenberger, 2000, S. 31ff., S. 37ff.; Böttcher u. a., 2008). Auf diese Weise kann die Signifikanz bestimmter Veränderungen bestimmt und mögliche Trends abgebildet werden (vgl. Rissland und Friedman, 1995). Da sich nicht jede Veränderung als gleich relevant erweist, gilt es aufzudecken, welche Veränderungen einen inhaltlichen Wert vorweisen und bei welchen es sich um vernachlässigbare Schwankungen handelt. Auf Basis der gemessenen relevanten Veränderungen kann dann ein Update des alten Musters erfolgen. Hierbei unterscheidet man zwei generelle Vorgehensweisen: *Strong Update* und *Weak Update* (vgl. Pechoucek u. a., 1999). Im ersten Fall wird die gesamte Struktur neu analysiert, die gemessenen Veränderungen werden lediglich zur Initialisierung der Analyse hinzugezogen. Ein Beispiel für dieses Vorgehen ist die bereits beschriebene Neubestimmung der Clusterzahl im Rahmen des Monitorings, die für die erneute Analyse verwendet werden kann (vgl. Angstenberger, 2000, S. 37ff.; Crespo und Weber, 2005). Bei der zweiten Vorgehensweise werden nur die Teilmuster neu berechnet, in denen eine signifikante Veränderung stattfand. So können beispielsweise beim Clustering lediglich diejenigen Cluster neu spezifizert werden, bei denen real eine Veränderung erfolgte. Letzteres Vorgehen birgt trotz seiner größeren Ungenauigkeit den Vorteil, dass es weniger komplex und damit nicht so kostenintensiv wie das vollständige Update des Gesamtstruktur ist. Die Komplexität der Change Mining-Verfahren darf nicht vernachlässigt werden: Bei immer wieder erfolgenden Veränderungen besteht die Möglichkeit, dass keine ausreichenden Ressourcen für ein ständiges vollständiges Update zur Verfügung stehen, so dass nur das Weak Update zur Anwendung kommt (vgl. Pechoucek u. a., 1999). Allgemein lässt sich jedoch festhalten, dass die Komplexität und die damit verbunden Kosten der Berechnungen aufgrund der Bedeutung aktueller Muster berechtigt sind (vgl. Chakrabarti u. a., 1998).

1.4. Bedeutung des Change Minings und der dynamischen Clusteranalyse für das Marketing

Wie in Abschnitt 1.3 einleitend beschrieben, setzt sich ein Unternehmen bei Nichtbeachtung der auftretenden Veränderungen über die Zeit hinweg einem hohen Risiko aus. Dies wird insbesondere dann zu bedrohlichen Folgen führen, wenn Veränderungen im Kundenverhalten analysiert oder beobachtete und aufgedeckte Veränderungen missachtet werden. Das Verständnis von Veränderungen sowie die Adaption der Marketingaktivitäten an diese erweist sich für das Überleben eines Unternehmens an einem sich verändernden Markt als elementar; besonders für internetbasierte Unternehmen wie Onlineshops ist die Analyse sich verändernder Daten aufgrund der Menge an erfassbaren Daten von Bedeutung (vgl. z.B. Li u. a., 2012; Song u. a., 2001). Werden Marketingentscheidungen auf Basis der auf veralteten Daten beruhenden Ergebnisse getroffen, können die Anstrengungen ihr Ziel verfehlen, da die Abstimmung auf das aktuelle Kundenverhalten und die Beachtung ihrer Präferenzen fehlt (vgl. Crespo und Weber, 2005). Als Konsequenz daraus folgt, dass mit einem aktiv betriebenen Change Mining und einer daraus resultierenden Adaption auftretender Umweltveränderungen ein erster Schritt zur Erzeugung eines Wettbewerbsvorteils durchgeführt wird (vgl. Wang u. a., 2003). Durch eine tiefergehende Datenanalyse im Sinne des Change Minings kann der Grundstein hierzu gelegt werden: Das Unternehmen erhält so einen Einblick in die Veränderungen, um sich strategisch daran auszurichten (vgl. Raghavan und Hafez, 2000). Die Anpassung an einen sich kontinuierlich verändernden Markt und an die variierenden Kundenbedürfnisse stellt somit für Unternehmen

eine Herausforderung dar (vgl. Chen u. a., 2005). Als Beispiel für verändertes Kundenverhalten kann der zunehmende Anteil von Internetnutzern in den späten 1990er und frühen 2000er Jahren betrachtet werden, durch den einerseits ein weiterer Distributionskanal geschaffen wurde und der andererseits im Bereich der Kommunikationspolitik das Online-Marketing initiierte. Unter anderem aufgrund der ökonomischen und marketingbezogenen Potentiale des Change Minings im Allgemeinen erfolgte die Entwicklung geeigneter dynamischer Techniken für verschiedene Data Mining-Verfahren zur Vorhersage des Kundenverhaltens; einige sollen in Kapitel 2 exemplarisch beschrieben werden.

Mit dem Einsatz der dynamischen Clusteranalyse in der Marktforschung gelingt es, durch Identifizieren aktueller Segmente und Analysieren ihrer Veränderungen gegenüber früheren Segmentierungsergebnissen ein aktuelleres Marktbild zu erhalten. Als Folge davon lassen sich unerwünschte Veränderungen, z.B. ein drohender Umsatzverluster in einzelnen Segmenten o.ä., frühzeitig aufdecken und geeignete Gegenmaßnahmen einleiten. Auch allgemeine Veränderungen im Kundenverhalten werden nachvollziehbar, z.B. die Beantwortung der Frage nach der Wirksamkeit bereits vorgenommener früherer Marketingaktionen oder dem Einfluss externer Faktoren. Damit bildet eine solche Analyse die Basis für erfolgreiche Marketingkampagnen (vgl. Böttcher u. a., 2008). In diesem Zusammenhang ist jedoch nicht nur die Beschreibung der Veränderungen, sondern auch die Vorhersage zukünftig möglicher Entwicklungen von hoher Relevanz, da einem erfolgreichen Bestehen am Markt das Erkennen und damit der Prognose zukünftigen Verhaltens elementare Bedeutung zukommt (vgl. u.a. Decker, 1994, S. 113f.; Sanz Sáiz, 2005). Erst das Aufdecken der Trends bzgl. Kundenverhalten und -präferenzen ermöglicht die Planung eines zukünftigen gezielten Einsatzes verschiedener Marketingsinstrumente: Im Rahmen der Produktpolitik können beispielsweise entsprechende Produkte entwickelt werden, die die Veränderung von Kundenpräferenzen einbeziehen. Ebenso können auch entstehende Nischen bedient werden, d.h. neu auftretende, wachsende Segmente, die durch die Clusteranalyse über die Zeit hinweg ermittelt werden. Ferner ist es möglich, das Verschwinden ganzer Kundensegmente frühzeitig aufzudecken, die in Zukunft im Rahmen eines differenzierten Marketings nicht weiter bedient werden müssen. Weitere Veränderungen in der Kundenstruktur insgesamt sind Segmente, die sich einander annähern und künftig als ein Segment betrachtet werden können oder aber verschiedene Segmente, die sich aus einem ursprünglich gemeinsamen Segment entwickeln. Im Marketingbereich ermöglicht die Kenntnis der genannten Veränderungen ein gezielteres Agieren. In der Kommunikationspolitik können so z.B. segmentspezifisch effektive Werbekampagnen geplant werden, insbesondere die Einbeziehung wahrscheinlicher Veränderungen innerhalb des Planungshorizonts ist möglich (vgl. z.B. Chen u. a., 2005; Song u. a., 2001).

1.5. Ausrichtung und Strukturierung der Arbeit

Das Ziel der vorliegenden Arbeit ist die Erweiterung der Fuzzy-Clusteranalyse auf dynamische Anwendungen. In diesem Zusammenhang erfolgt eine Anpassung bereits bekannter Ansätze zur possibilistischen Clusteranalyse, um diese in einem sich verändernden Umfeld einsetzen zu können. Ferner werden Möglichkeiten evaluiert und ergänzt, so dass eine Analyse von Veränderungen einer Clusterstruktur im Sinne des Change Minings vorgenommen werden kann, und Ansätze zum Erkennen zukünftiger Entwicklungen werden geliefert; darin liegt auch die Erweiterung bereits bekannter Ansätze. Sowohl Angstenberger (2000) als auch Crespo und Weber

(2005) bzw. Weber (2007) untersuchen Veränderungen innerhalb einer Clusterstruktur, jedoch handelt es sich dabei um ausschließlich deskriptive Modelle. Crespo und Weber beschränken sich auf das probabilistische Clustering (vgl. Abschnitt 3.3; Crespo und Weber (2005); Weber (2007)), welches nur bedingt für die Analyse im Sinne des Change Minings geeignet ist; da sie sich jedoch auf die Bestimmung der Clusterzahl für Folgeperioden beschränken, ist diese Einschränkung gerechtfertigt. Sollen hingegen konkrete Entwicklungen innerhalb der Clusterstruktur nachvollzogen sowie zukünftige Anpassungen vorhergesagt werden, reicht dieser Ansatz nicht länger aus. Angstenberger (2000) untersucht keine sich ändernden Objekte, sondern Funktionen und Szenarien. Dabei liegt ihr Fokus auf der Ermittlung von Unterschieden gegenüber bisheriger Modelle; Ziel der Analyse ist auch hier in erster Linie das Ermitteln einer geeigneten Clusterzahl für eine erneute Clusteranalyse. Dabei werden verschiedene Änderungstypen wie das Neubilden von Clustern sowie das Vereinigen und das Trennen einzelner Cluster näher betrachtet, die durch die erkannten graduellen Änderungen implizierten Entwicklungen sind jedoch nicht Teil der Untersuchung. Die in dieser Arbeit dargestellte Vorgehensweise bietet neben der Bestimmung bereits aufgetretener Veränderungen außerdem die Möglichlichkeit, zukünftige Veränderungen vorherzusagen. Dies ist besonders im Rahmen des Change Minings im Marketing elementar, um auf anstehende Veränderungen in ausreichendem Maße reagieren zu können (vgl. Kapitel 5). Bei dem beschriebenen Ansatz handelt es sich demnach nicht um ein rein deskriptives, sondern um ein prädiktives Modell.

Die Arbeit ist wie folgt strukturiert: Zunächst wird die dynamische Analyse unter Anwendung bestimmter Data Mining-Verfahren näher betrachtet. Hierbei wird insbesondere auf die für das Marketing relevanten Verfahren zur Multidimensionalen Skalierung und zum Association Rule Mining eingegangen (Kapitel 2). In Kapitel 3 werden anschließend die grundsätzlichen Aspekte des Fuzzy-Clusterings näher erläutert. Kapitel 4 und 5 stellen den eigentlichen Kern der Arbeit dar. In Kapitel 4 werden die Erweiterungsmöglichkeiten zum possibilistischen Fuzzy-Clustering vorgestellt, um die Anwendung im Sinne des Change Minings zu verbessern. Kapitel 5 befasst sich mit dem tatsächlichen Prozess des Change Minings bei der Fuzzy-Clusteranalyse. Es werden mögliche Veränderungen aufgezeigt und erläutert, wie diese ermittelt und gegebenenfalls prognostiziert werden können. Abschließend erfolgt in Kapitel 6 eine Schlussbetrachtung sowie ein Ausblick auf weitere Forschungsfragen.

If you torture the data long enough, Nature will confess.

R. Cease

2

Dynamische Analyse unter Anwendung verschiedener Data Mining-Verfahren

Das Change Mining verfügt in verschiedenen Gebieten über eine hohe Bedeutung, so dass für die einzelnen Bereiche des Data Minings Ansätze zum Change Mining entwickelt wurden (vgl. Crespo und Weber, 2005; Weber, 2007). So gibt es zur Aktualisierung von Entscheidungsbäumen verschiedene Techniken zur Baumrekonstruktion sowie zum inkrementellen und induktiven Lernen auf Basis bereits bekannten Wissens (vgl. Pechoucek u. a., 1999). In diesem Zusammenhang ist die enge Verwandtschaft zum hierarchischen Clustering zu beachten, bei dem das Ergebnis in einem Dendrogramm dargestellt wird; auch hier kann eine Anpassung über inkrementelles Lernen erfolgen (vgl. u.a. Kalnis u. a., 2005; Nassar u. a., 2004). Einige der Ansätze zur Nachverfolgung von Veränderungen, die insbesondere im Bereich der Marktforschung eine hohe Relevanz besitzen, sollen im Folgenden näher betrachtet werden: die Multidimensionale Skalierung als Möglichkeit zur Darstellung von Marktsegmenten sowie das Aufdecken von Assoziationsregeln (Association Rule Mining) zum Verdeutlichen von Kaufverhaltensmustern.

Für die dynamische Analyse von Assoziationsregeln werden in der Literatur unterschiedlichste Systeme vorgestellt, die aufgrund der bestehenden Vielfalt im Zusammenhang mit dem Change Mining in Abschnitt 2.2 ausführlich betrachtet werden. Vorab soll zunächst jedoch ein klassisches Data Mining-Verfahren detailliert vorgestellt werden, das in der Marktforschung im Rahmen der Interdependenzanalyse über hohe Bedeutung verfügt: die Multidimensionale Skalierung (MDS).

2.1. Multidimensionale Skalierung (MDS)

Die Multidimensionale Skalierung (MDS), auch Mehrdimensionale Skalierung genannt, verfolgt das Ziel, eine Repräsentation von (Un-)Ähnlichkeitsmessungen zwischen Objektpaaren als Distanzen im niedrigdimensionierten Raum zu erstellen (vgl. Borg und Groenen, 2005, S. 3) und so eine visuelle Untersuchung der Objektstruktur zu ermöglichen. Als Begründer der MDS gelten Shepard (1957), der 1957 in der Psychometrika seinen ersten Artikel zum Thema veröffentlichte, sowie Kruskal (1964), der 1964 eine weiterentwickelte Form der MDS vorstellte. Der Ursprung der MDS ist demnach in der Psychologie zu finden.

2.1.1. Einführung Multidimensionale Skalierung

Die MDS wird angewandt, um den Wahrnehmungsraum einer Untersuchungsperson bzgl. bestimmter Objekte mit Hilfe mathematischer Modelle zu visualisieren und damit eine räumliche Anordnung der Objekte im m-dimensionalen Raum zu ermöglichen. Im Laufe der Jahre hielt sie Einzug in unterschiedlichste Bereiche; als Beispiel hierfür sind unter anderem das Human Resource Management oder die Logistik zu nennen (vgl. Neumann, 2006, S. 8). Zu ihren weiteren Anwendungsgebieten gehört im Rahmen des Marketings die Marktforschung, wo ihre Bedeutung darin besteht, den subjektiven Wahrnehmungsraum potentieller Konsumenten von Produkten oder Dienstleistungen zu bestimmen. Auf diese Weise ist es unter anderem möglich, folgende Fragestellungen zu beantworten:

- Welche Produkte sind aus Kundensicht substituierbar?
- Wo befinden sich in der Käuferwahrnehmung Marktnischen bzw. Marktlücken?
- Welche Produkte kommen der Idealvorstellung der Nachfrager am nächsten?
- Welche Eigenschaften sind für die Beurteilungen der Ähnlichkeiten relevant?

Die MDS bietet somit eine Möglichkeit, Markt- oder Produktsegmente grafisch zu veranschaulichen und ggf. Veränderungen über einzelne Perioden hinweg zu verdeutlichen.

Der Ausgangspunkt der klassischen MDS basiert auf der Annahme, dass alle Untersuchungspersonen über einen nahezu identischen Wahrnehmungsraum verfügen. Damit ermöglicht sie einem Unternehmen, unter anderem die Attraktivität der eigenen Produkte zu ermitteln oder entscheidungsrelevante Eigenschaften zu identifizieren. In diesem Fall werden die wahrgenommenen globalen Ähnlichkeiten in der Untersuchungsgruppe erhoben und zu einer gemeinsamen Ähnlichkeitsmatrix aggregiert, die anschließend in eine Distanzmatrix transformiert wird; es folgt die Bestimmung einer einheitlichen Konfiguration. Dabei unterscheidet man zwei wesentliche Arten der Multidimensionalen Skalierung: die metrische MDS nach Shepard und die nichtmetrische nach Kruskal. Im Fall der metrischen MDS sollen die Zieldistanzen (Disparitäten) in der resultierenden Konfiguration den erhobenen Distanzen entsprechen. Im praktischen Anwendungsbereich ist eine solche Lösung jedoch i.d.R. nicht im Bereich des Möglichen; dies liegt unter anderem darin begründet, dass viele Objekte im niedrigdimensionierten Raum abgebildet werden sollen. Zudem stammen die Daten aus aggregierten Konsumentenbefragungen und sind somit oftmals widersprüchlich (vgl. Neumann, 2006, S. 15f.). Ferner fehlt die Gewähr, dass die erhobenen Distanzen die Wahrnehmung der Konsumenten exakt widerspiegeln, da zur Berechnung der Distanzen im Allgemeinen Ähnlichkeiten herangezogen werden, deren Erhebung nicht selten in Form von Ordinaldaten erfolgt. Aus den genannten Gründen ist die Anwendung der metrischen MDS wenig erfolgsversprechend.

Im Gegensatz zur metrischen MDS werden bei der nichtmetrischen MDS vergleichsweise wenige Prämissen verlangt, meist aber gleichwertige Ergebnisse erreicht (vgl. Schobert und Dichtl, 1979, S. 9). Als Basis für die nichtmetrische MDS dient die Bedingung, dass die Rangreihung der Proximitäten, d.h. die Annäherung der tatsächlichen Distanzen in der Konfiguration, mit der Rangreihung der erhobenen Distanzen übereinstimmt; die Exaktheit der Distanzen wird nicht verlangt. Diese Monotoniebedingung ist bei n Objekten im $(n-1)$-dimensionalen Raum stets erfüllbar (vgl. Green u. a., 1989, S. 42). Im Allgemeinen wird diese Form der MDS den metrischen Verfahren vorgezogen, da die Ergebnisse nicht weniger genau, die resultierenden

Konfigurationen jedoch wesentlich robuster in Bezug auf Messungenauigkeiten o.ä. sind (vgl. Schobert und Dichtl, 1979, S. 9f.). Dieser Umstand folgt nicht zuletzt aus der Tatsache, dass bei der nichtmetrischen MDS in einem zweistufigen Iterationsprozess die Optimierung des Gütekriteriums erfolgt, genannt *Stress*.

Definition 2.1. Die quadratische Abweichung zwischen den Proximitäten $\hat{d}(\vec{x}_j, \vec{x}_k)$, d.h. der Distanz innerhalb einer Konfiguration zwischen zwei Objekten j und k, und der Monotonieanpassung $\delta(\vec{x}_j, \vec{x}_k)$, die die Rangreihung der erhobenen Distanzen widerspiegelt, wird als *Rohstress* bezeichnet:

$$b(\vec{x}_1, \ldots, \vec{x}_n) = \sum_{\substack{j,k=1 \\ j<k}}^{n} \left(\hat{d}(\vec{x}_j, \vec{x}_k) - \delta(\vec{x}_j, \vec{x}_k)\right)^2$$

Durch Normierung des Rohstresses, z.B. über die Division mit dem maximal möglichen Wert $b_{\max}$, erhält man einen mit anderen Konfigurationen vergleichbaren Wert, den *Stress*.

Um die Prämisse der klassischen MDS, dass alle Untersuchungspersonen über denselben Wahrnehmungsraum verfügen, zu umgehen, entwickelten Carroll und Chang (1970) darauf aufbauend den sogenannten INDSCAL-Algorithmus (*In*dividual *D*ifferences *Scal*ing), der das Ziel verfolgt, eine einzige räumliche Darstellung der Objekte zu erstellen. Um die individuellen Beurteilungen der Ähnlichkeiten zu berücksichtigen, wird eine Gruppenkonfiguration bestimmt, der sogenannte *group stimulus space* (vgl. Borg und Groenen, 2005, S. 474), die entlang ihrer Dimensionen gestreckt oder gestaucht werden kann, um die individuellen Konfigurationen zu erhalten. Dazu werden individuelle Gewichte je Dimension und Subjekt vergeben, die als Bedeutungsgrad der Dimensionen für das jeweilige Subjekt interpretiert werden können. Die Bestimmung der Gewichte erfolgt im Verlauf des Verfahrens, die der individuellen Konfigurationen unterbleibt, da sie sich im Anschluss mit Hilfe der Gesamtkonfigurationen und der individuellen Gewichte unproblematisch herleiten lassen. Um dieses Vorgehen zu ermöglichen, werden die Distanzen unter Anwendung der gewichteten euklidischen Metrik bestimmt, so dass die Dimensionsgewichte auch in der Bestimmung des Stresswertes eingehen können. Damit verändert sich die in Definition 2.1 gegebene Stressformel bei G individuellen Konfigurationen zu

$$b_{\text{INDSCAL}}(X_1, \ldots, X_G) = \sum_{g=1}^{G} \sum_{\substack{j,k=1 \\ j<k}}^{n} \left(d_{jk}^g - \delta^g(\vec{x}_j, \vec{x}_k)\right)^2$$

Dabei werden in der Regel keine weiteren Transformationen der erhobenen Distanzen d_{jk}^g vorgenommen, vielmehr werden diese direkt mit den zugehörigen Monotonieanpassungen in der g-ten Konfiguration $\delta^g(\vec{x}_j, \vec{x}_k)$ verglichen (vgl. Neumann, 2006, S. 39).

Anhand der Ergebnisse des Algorithmus lassen sich diejenigen Dimensionen ermitteln, die allgemein für alle Subjekte die Basis ihrer Ähnlichkeitsurteile bilden; die erhaltenen Ergebnisse bieten ferner Erkenntnisse über individuell relevante Dimensionen. Sie sind daher im Bereich der Marktforschung von besonderem Interesse, da sie hier die Unterscheidung der für einzelne Konsumentengruppen relevanten Aspekte im Gegensatz zu den bedeutungslosen und am Ende unter Umständen herausfallenden Entscheidungskriterien ermöglichen.

2.1.2. Ansätze zum Change Mining

Im Rahmen der Marktforschung erfordern zwei wesentliche Überlegungen die dynamische Analyse im Zusammenhang mit der Multidimensionalen Skalierung: Zum einen kann durch eine solche Betrachtung die Wirkung von Marketingaktivtäten und die daraus resultierenden Veränderungen in der Einstellung der Konsumenten bzgl. der untersuchten Objekte analysiert werden, zum anderen können die Objekte selbst im Laufe der Zeit verändert worden sein. In beiden Fällen ist es für die Marktforschung von elementarer Bedeutung, Trends aufzudecken; so können sich z.B. die Relationen zwischen den Objekten im Untersuchungszeitraum ändern. Das Ziel einer entsprechenden Analyse liegt sowohl im Hervorheben endogener sowie im Aufdecken exogener Einflüsse, die als Ursache für die Veränderungen herangezogen werden können (vgl. Schobert, 1979, S. 90).

Intuitiv erscheint es naheliegend, Konfigurationen im niedrigdimensionierten Raum rein optisch zu vergleichen, um mögliche Unterschiede zu erkennen. Dieses einfache Vorgehen erweist sich jedoch als problematisch, da für die optisch wahrnehmbaren Differenzen der Konfigurationen verschiedene Ursachen verantwortlich sein können. Eine mögliche Ursache liegt neben den allgemein vorhandenen Erhebungsungenauigkeiten in methodischen Verzerrungen, die aus unterschiedlichen Erhebungsmethoden während der Konsumentenbefragung resultieren. So können z.B. in einer Periode die Objekte mittels Rangreihung verglichen, in der Folgeperiode die Ähnlichkeiten über ein Rating-Verfahren gemessen werden; auch können methodische Unterschiede bei der Erstellung der Konfigurationen dafür verantwortlich sein. Weiterhin können bestimmte konfigurationsbedingte Veränderungen nur schwer von tatsächlichen datenbezogenen Differenzen unterschieden werden. Ein oberflächlicher optischer Vergleich ist besonders bei einer großen Objektzahl kaum möglich, insbesondere, wenn eine unterschiedliche räumliche Orientierung der Objekte in den einzelnen Konfigurationen vorliegt (vgl. Neumann, 2006, S. 51f.). So würde z.B. eine einfache Rotation oder Spiegelung einer Konfiguration ein optisch stark unterschiedliches Ergebnis generieren, inhaltlich bliebe die Konfiguration jedoch unverändert. Die Beantwortung der grundsätzlichen Frage, inwieweit sich zwei Konfiguration *wirklich* voneinander unterscheiden, definiert das durch die dynamische MDS zu erreichende Ziel. Ihre Aufgabe besteht darin, unrelevante Differenzen von den relevanten zu separieren; dies ist in höherdimensionierten Räumen kaum möglich, aber bereits die zweidimensionale Darstellung ist i.d.R. sehr hilfreich, um die resultierenden Unterschiede hervorzuheben (vgl. Borg und Groenen, 2005, S. 64ff.).

Bei der Dynamisierung der MDS können drei Arten unterschieden werden: die parallele, die simultane und die sequentielle Dynamisierung, die im Folgenden näher erläutert werden. Jedes dieser Dynamisierungsmodelle geht dabei von denselben Prämissen aus (vgl. Neumann, 2006, S. 52):

- Für die T betrachteten Perioden sind die zugehörigen $n \times n$-Distanzmatrizen D_t, $t = 1, \ldots, T$ gegeben.
- Der betrachtete Zeitraum besteht aus mindestens zwei Perioden ($T \geq 2$).

Parallele Dynamisierung

Bei der parallelen Dynamisierung werden alle T Konfigurationen $X_t, t = 1, \ldots, T$, parallel, d.h. nebenläufig erstellt; damit entfällt eine anschließende Anpassung der Konfigurationen. Schobert (1979, S. 224ff.) schlägt vor, die Dynamisierung mit Hilfe des bereits bekannten INDSCAL-Algorithmus (vgl. Abschnitt 2.1.1) durchzuführen, indem anstelle der G Distanzmatrizen $D_g, g = 1, \ldots, G$, periodenspezifische Distanzmatrizen $D_t, t = 1, \ldots, T$, verwendet werden. Auf diese Weise erfolgt die Bestimmung einer gemeinsamen Basis aller Perioden; die periodenspezifischen Konfigurationen X_t sind über die resultierenden Dimensionsgewichte definiert. Dieses Vorgehen bietet den Vorteil, dass für jede Periode ein separater Modellraum existiert, der gesondert analysiert und evaluiert werden kann. Der Zusammenhang zwischen den einzelnen Perioden wird durch die Gewichtungsmatrix gegeben, die auch bereits erste Anstöße für eine mögliche Interpretation der Veränderung birgt. So könnte sich z.B. ein Dimensionsgewicht über den untersuchten Zeitraum merklich erhöhen; eine solche Veränderung spräche dann für eine wachsende Relevanz dieser Dimension in der Konsumentenwahrnehmung. Durch die Verwendung des INDSCAL-Algorithmus ist dieses Vorgehen sehr einfach zu handhaben, da die Bestimmung der einzelnen Konfigurationen nur einen Arbeitsschritt erfordert. Ferner lässt sich dieser Ansatz problemlos auf ein individuelles dynamisches Modell erweitern, da eine Unterscheidung zwischen individuellen oder periodenspezifischen Distanzmatrizen bei der Durchführung der eigentlichen Analyse nicht erfolgt.

Ein wesentlicher Nachteil der parallelen Dynamisierung besteht jedoch in der fehlenden Zuverlässigkeit des INDSCAL-Algorithmus beim Vorkommen von Missing Values, so dass die betrachtete Objektmenge im Beobachtungszeitraum nicht variieren darf (vgl. Neumann, 2006, S. 55). Besonders im Rahmen der Marktforschung und des Marketings erweist sich diese Bedingung als undurchführbar, da immer wieder neue Produkte eingeführt und alte eliminiert werden. Weiterhin wird durch die Berechnung einer gemeinsamen Gruppenkonfiguration ein Zusammenhang konstruiert, der in der Realität nicht zwangsläufig besteht; diese Tatsache erzeugt eine starke Reduzierung der Sensitivität gegenüber Periodenschwankungen. So ist es möglich, dass ein untersuchtes Objekt zwischen zwei Perioden stark an Ansehen in der relevanten Untersuchungsgruppe verliert, z.B. durch einen im zugehörigen Unternehmen eingetretenen Krisenfall, die übrigen Objekte davon jedoch nahezu unberührt sind. Ein weiterer, für das Change Mining weitaus schwerwiegender Nachteil besteht zudem darin, dass das Hinzufügen einer weiteren Periode im Zeitverlauf die Neuberechnung des vollständigen Modells erfordert, da es sich als unmöglich darstellt, die Veränderungen schrittweise bzw. über die Betrachtung einzelner Zeitfenster zu analysieren.

Simultane Dynamisierung

Bei der simultanen Dynamisierung erfolgt eine gemeinsame Verarbeitung der einzelnen T Distanzmatrizen durch eine Aggregation dieser zu einer gemeinsamen Superdistanzmatrix D^S mit

$$D^S = \begin{pmatrix} D_{11} & \ldots & D_{1T} \\ \vdots & \ddots & \vdots \\ D_{T1} & \ldots & D_{TT} \end{pmatrix}, \text{ wobei } D_{tt'} = \begin{pmatrix} d_{11}^{tt'} & \ldots & d_{1n}^{tt'} \\ \vdots & \ddots & \vdots \\ d_{n1}^{tt'} & \ldots & d_{nn}^{tt'} \end{pmatrix}$$

(vgl. Schobert, 1979, S. 229). Die zwischenperiodischen Distanzmatrizen $D_{tt'}$ sind im Allgemeinen nicht symmetrisch, da die Distanz $d_{jk}^{tt'}$ die Distanz zwischen Objekt j zum Zeitpunkt t und Objekt k zum Zeitpunkt t', $d_{kj}^{tt'}$ hingegen die Distanz zwischen Objekt k zum Zeitpunkt t und Objekt j zum Zeitpunkg t' angibt. Ebenso gilt nicht zwangsläufig $d_{jj}^{tt'} = 0$ für $t \neq t'$, da sich ein Objekt zwischen den Perioden verändert haben kann. Bei diesem Vorgehen werden keine separaten periodenspezifischen Konfigurationen erstellt, die entsprechenden Resultate lassen sich jedoch problemlos periodenspezifisch aus der Gesamtkonfiguration herauslösen. Unter der Prämisse, dass die Superdistanzmatrix D^S vollständig bekannt ist, birgt dieser Ansatz mehrere Vorteile. So kann im Gegensatz zur parallelen Dynamisierung mittels INDSCAL die Objektmenge ohne Einfluss auf die Verarbeitung variieren. Ferner wird keine gemeinsame Gruppenkonfiguration konstruiert, so dass individuelle Periodenunterschiede deutlich hervorgehoben werden und die Gefahr von Periodenfehlern deutlich verringert wird. Auch ist hier keinerlei anschließende Anpassung erforderlich, da die Beziehungen aller Objekte zu den unterschiedlichen Zeitpunkten zueinander dargestellt werden.

Als nachteilig erweist sich, dass die einfache Kenntnis der periodenspezifischen Distanzmatrizen D_{tt} nicht ausreicht; zudem werden Vergleiche zwischen den Perioden benötigt. Diese Möglichkeit besteht, wenn für die Analyse über den Beobachtungszeitraum sowohl in Merkmalsmenge als auch im Messverfahren bei der Erhebung vergleichbare Distanzen verwendet werden, die auf Profildaten anstelle von erhobenen Ähnlichkeitswahrnehmungen basieren. Zudem steigt der Rechenaufwand überproportional zum Umfang der Objektmenge. Die hohe Objektzahl von $n \times T$ Objekten zeigt dabei einen ausgeprägten Einfluss auf die Dimensionierung der resultierenden Konfiguration und damit eine deutliche Tendenz zu höher dimensionierten Räumen. Des Weiteren erweist sich dieses Vorgehen zur Anwendung in der Marktforschung und damit der Ermittlung subjektiver Wahrnehmungsräume als ungeeignet, wenn im Bereich der Marktforschung meist nur die periodenspezifische Distanzmatrizen bekannt sind. Ferner besteht aufgrund der Vielzahl an Missing Values eine starke Neigung der MDS zur Darstellung einzelner Teilgraphen, ohne dass eine Beziehung zwischen den einzelnen Perioden hergestellt werden kann (vgl. Schobert, 1979, S. 229f.).

Um das Problem der Notwendigkeit einer vollständigen Superdistanzmatrix zu umgehen, entwickelten Ambrosi und Hansohm (1986) einen Ansatz zur simultanen Dynamisierung, der die Anpassung der einzelnen Konfigurationen aneinander simultan zur Erstellung der Gesamtkonfiguration vornimmt und dabei nur die periodeninternen Distanzmatrizen $D_{tt} = D_t$ einbezieht. Dies erfolgt mit Hilfe eines Optimierungsproblems:

$$h_\phi\left(D_1, \ldots, D_T, X_1, \ldots, X_T\right) = b_{sim}\left(D_1, \ldots, D_T, X_1, \ldots, X_T\right) + \phi \cdot a\left(X_1, \ldots, X_T\right) \rightarrow \min$$

wobei

- $b_{sim}\left(D_1, \ldots, D_T, X_1, \ldots, X_T\right) = \sqrt{\dfrac{\sum_{t=1}^{T} \sum_{j<k} \left(d_{jk}^t - \delta^t\left(\vec{x}_j, \vec{x}_k\right)\right)}{\sum_{t=1}^{T} \sum_{j<k} \left(d_{jk}^t\right)^2}}$: Stressfunktion
- $a\left(X_1, \ldots, X_T\right) = \sum_{t=1}^{T-1} \sum_{j=1}^{n} \sum_{l=1}^{p} \left(x_{jl}^{t+1} - x_{jl}^t\right)^2$: Anpassungsfunktion für p Dimensionen

- $\phi > 0$: Gewichtungsfaktor für den Einbeziehungsgrad der Anpassung

Auf diese Weise lassen sich die ermittelten Konfigurationen ohne weiteren Anpassungsschritt in einem gemeinsamen Raum darstellen. Der entscheidene Vorteil gegenüber der Verwendung einer Superdistanzmatrix besteht darin, dass auch bei alleiniger Kenntnis der periodenspezifischen Distanzmatrizen D_t eine simultane Dynamisierung möglich bleibt und ein ausreichend gutes, zusammenhängendes Ergebnis erzielt werden kann. Dessen ungeachtet bleibt beiden vorgestellten Beispielen der simultanen Dynamisierung, dass analog zur parallelen Dynamisierung alle Konfigurationen in einem Schritt erstellt werden und als Folge davon im fortschreitenden Zeitverlauf neue Perioden nur durch eine vollständig neu durchgeführte Analyse aufgenommen werden können.

Sequentielle Dynamisierung

Im Rahmen der sequentiellen Dynamisierung werden die einzelnen Dynamisierungsschritte nacheinander ausgeführt. Zunächst erfolgt eine separate Bestimmung der T Konfigurationen $X_t, t = 1, \ldots, T$; im Anschluss werden diese mit Hilfe einer generalisierten Prokrustes-Analyse aneinander angepasst. In diesem Schritt wird versucht, die T Konfigurationen X_t optimal aufeinander abzubilden, um konfigurationsbedingte Differenzen zu eliminieren. Es handelt sich dabei um ein Drei-Wege-Modell, da von n Objekten, p Dimensionen sowie T Konfigurationen ausgegangen wird (vgl. Borg und Groenen, 2005, S. 57f.). Als zulässige Prokrustes-Transformationen[3] werden solche bezeichnet, die keinen Einfluss auf das Verhältnis der Distanzen innerhalb einer Konfiguration zueinander haben, d.h. Strecken bzw. Stauchen der Konfiguration, Verschieben und Rotieren der Konfiguration sind erlaubt[4]. In der Simplizität dieses Vorgehens liegt sein Nachteil: Die anschließende Anpassung der Konfigurationen aneinander anstatt gleich bei ihrer Erstellung erfordert zusätzlichen Aufwand, der speziell dann erhöht wird, wenn die Anpassung mittels der generalisierten Prokrustes-Analyse simultan erfolgt, d.h., wenn alle Konfigurationen gleichzeitig einander angepasst werden. Des Weiteren werden allgemeine Trends eventuell vernachlässigt, ein Nachteil gegenüber der simultanen Dynamisierung mittels einer Superdistanzmatrix auf Basis von Profildaten, die diese darstellen kann.

Als vorteilhaft ergibt sich die Möglichkeit, für jede Periode eine eigene separate Konfiguration erstellen zu können, ohne einen konstruierten Zusammenhang wie bei der parallelen Dynamisierung zugrunde zu legen. Auf diese Weise können Extremallösungen und starke Schwankungen zwischen den Perioden verdeutlicht werden; dies ist besonders dann von Bedeutung, wenn der Effekt bestimmter Marketingmaßnahmen gemessen werden soll. Wird die Objektzahl variiert, muss bei der Prokrustesanalyse jedoch die Relevanz der Objekte für die Anpassung beachtet werden. Damit bietet die sequentielle Dynamisierung ein breiteres Einsatzspektrum, da hier die Objektmenge nachträglich variiert werden kann. Für das Change Mining als solches ist jedoch entscheidend, dass für jede neu hinzukommende Periode bzw. für das jeweils betrachtete Zeitfenster eine eigenständige Konfiguration erstellt werden kann, ohne dass die bereits vorhandenen Konfigurationen angepasst werden müssen. Diese neue Konfiguration wird anschließend

[3] Der Name *Prokrustes* ist der griechischen Mythologie entnommen. Prokrustes passte seine Gäste der Größe ihres Bettes an, indem er sie entweder langzog oder ihre Beine verkürzte (vgl. Mathar, 1997, S. 33).

[4] Genauere Informationen zur einfachen Prokrustes-Analyse sind Neumann (2006, S. 30ff.), zum Drei-Wege-Prokrustes-Modell Gower und Dijksterhuis (2004) zu entnehmen.

mit Hilfe einer einfachen Prokrustes-Transformation an die vorherigen Perioden angepasst, so dass auftretende Veränderungen analysiert und evaluiert werden können.

Unabhängig von der Art der Dynamisierung kann neben einer einfachen Skalierung die Interpretierbarkeit auch bei einer dynamischen Analyse analog zur klassischen statischen MDS durch das Einfügen von Eigenschaftsvektoren verbessert werden (vgl. Neumann, 2006, S. 24ff.). In diesem Zusammenhang bietet die sequentielle Dynamisierung den Vorteil, dass zusätzlich die Möglichkeit einer Analyse der Veränderung der Bedeutung von Eigenschaften besteht, die ihrerseits den Wahrnehmungsraum der Konsumenten prägen und im Untersuchungszeitraum einer Veränderung unterliegen. So können vorhandene Eigenschaften durchaus an Bedeutung verlieren, während andere in der Kundenwahrnehmung deutlichen Zuwachs erfahren; ein Beispiel hierfür stellt die wachsende Sensibilität gegenüber umweltbezogenen Aspekte dar, die es ihrerseits vermag, die Wahrnehmung eines Kunden zu beeinflussen.

2.2. Dynamisches Association Rule Mining

Das Aufdecken von Assoziationsregeln, das sogenannte *Association Rule Mining*, wurde in den 1990er Jahren erstmalig eingeführt; im Bereich des unüberwachten Lernens gehört es bis heute zu den bedeutenden Forschungsgebieten (vgl. Agrawal u. a., 1993; Raghavan und Hafez, 2000). Das zugrundeliegende Problem ist auch unter dem Namen *market basket problem* bekannt.

2.2.1. Einführung Association Rule Mining

Das Association Rule Mining befasst sich mit der Analyse von Kundenverhaltensmustern, auf deren Basis gezielt Produkte angeboten und platziert werden können; es gilt also, Abhängigkeiten zwischen Attributwerten in der Form *Kunden, die A kaufen, kaufen auch B* aufzudecken (vgl. Cios u. a., 2007, S.290f.). Ein Beispiel für die Anwendung der Assoziationsregeln sind Recommender, die Kunden anhand der erlernten Verhaltensweisen weitere Produkte empfehlen.

Formal lassen sich Assoziationsregeln wie folgt definieren (vgl. Raghavan und Hafez, 2000):

Definition 2.2. Sei $\mathfrak{I} = \{\iota_1, \ldots, \iota_w\}$ eine Menge an Items und $\mathfrak{T} = \{\psi_1, \ldots, \psi_n\}$ die Menge an durchgeführten Transaktionen, wobei gilt $\forall \psi_j \in \mathfrak{T} : \psi_j \subseteq \mathfrak{I}$. Dann ist $A \Rightarrow B$ eine *Assoziationsregel*, sofern gilt $A, B \subset \mathfrak{I} \wedge A \cap B = \emptyset$.

Assoziationsregeln beschreiben die Implikation von Itemmenge A zu Itemmenge B; dabei sind die einzelnen Regeln von unterschiedlicher Bedeutung. Zunächst muss die Stärke der Beziehung zwischen den Itemmengen mittels verschiedener probabilistischer Maße evaluiert werden (vgl. Bramer, 2007, S. 190; Raghavan und Hafez, 2000):

1. *Support*: Anteil an Transaktionen in $\mathfrak{T}$, die sowohl A als auch B enthalten:

$$Support(A \Rightarrow B) = P(A, B)$$

2. *Confidence*[5]: Anteil an Transaktionen in $\mathfrak{T}_A$, die B enthalten, wobei $\mathfrak{T}_A$ die Menge aller Transaktionen darstellt, die A enthalten. Die Confidence gibt somit den Anteil korrekter Vorhersagen von B basierend auf A an:

$$Confidence(A \Rightarrow B) = \frac{P(A,B)}{P(A)}$$

3. *Completeness*: Anteil der durch die Regel generierten korrekten Vorhersagen von B basierend auf allen Vorkommnissen von B:

$$Completeness(A \Rightarrow B) = \frac{P(A,B)}{P(B)}$$

4. *Interest*: Test für statistische Unabhängigkeit zwischen A und B:

$$Interest(A \Rightarrow B) = \frac{P(A,B)}{P(A)P(B)}$$

Um relevante Assoziationsregeln zu explorieren, werden zunächst sogenannte *häufige Itemmengen* ermittelt.

Definition 2.3. Sei $A \subset \mathfrak{I}$ eine Itemmenge. Falls gilt $Support(A) = P(A) \geq Support_{\min}$, wobei $Support_{\min}$ den minimalen Support angibt, ist A eine *häufige Itemmenge* (vgl. Raghavan und Hafez, 2000).

Zur Ermittlung aller häufigen Itemmengen können alle möglichen Kombinationen durchgeführt und ihr Support überprüft werden. Dieses naive Vorgehen erfordert jedoch hohen Aufwand ($O(2^w n)$) (vgl. Cios u. a., 2007, S. 295); die Anwendung des von Agrawal und Srikant (1994) eingeführten Standardalgorithmus des Association Rule Mining, der Apriori-Algorithmus, erweist sich als effizienter. Wie der Name bereits impliziert, wird in diesem Algorithmus zur Beschleunigung bereits bekanntes Wissen, das Apriori-Wissen, genutzt: Alle nichtleeren Teilmengen einer häufigen Itemmenge müssen nach Definition 2.3 ebenfalls häufig sein. Im Umkehrschluss bedeutet dies, dass bei fehlender Häufigkeit einer gegebenen Itemmenge auch keine ihrer Obermengen häufig vorkommt[6]. Der Apriori-Algorithmus beginnt daher mit einelementigen Itemmengen, kombiniert diese dann zu möglichen zweielementigen etc.

Aus den so ermittelten häufigen Itemmengen werden die relevanten Assoziationsregeln nach der folgenden Regel generiert (vgl. Cios u. a., 2007, S. 297):

Definition 2.4. Sei B eine häufige Itemmenge. Dann ist $A \Rightarrow B \setminus A, A \subset B$ eine *relevante Assoziationsregel* r, sofern gilt $Confidence(A \Rightarrow B \setminus A) \geq Confidence_{\min}$, wobei $Confidence_{\min}$ die minimale Confidence angibt.

[5] weitere Bezeichnungen: *Predictive Accuracy*, *Reliability*

[6] Für den vollständigen (intuitiven) Beweis siehe z.B. Cios u. a. (2007, S. 295).

2.2.2. Ansätze zum Change Mining

Assoziationsregeln sind nicht statisch, sondern können sich über die Zeit hinweg verändern. So besteht einerseits die Möglichkeit, dass Regeln, die bestimmte Zusammenhänge beschreiben, nicht mehr gelten. Andererseits kommt es zum Auftreten neuer Regeln, weil z.B. neue Kombinationsmöglichkeiten bestimmter Produkte propagiert werden oder aber – im Falle von Assoziationsregeln, die von Kundeneigenschaften auf Kaufentscheidungen schließen lassen – weil Produkte auch von anderen Käufergruppen adaptiert werden.

Erste Ansätze zum Change Mining im Gebiet der Assoziationsregeln liefern Agrawal und Psaila (1995). Die Autoren versuchen, mit Hilfe eines kontinuierlichen Minings mit vorgegebener Frequenz Regeln aufzudecken; dabei wird zunächst eine wiederholte Analyse ohne Einbeziehung von Apriori-Wissen aus den Vorperioden durchgeführt. Die entdeckten Assoziationsregeln können in zwei Klassen unterteilt werden: bereits bekannte Regeln und neu erlernte Regeln. Letztere werden lediglich der Regelbasis R hinzugefügt, so dass sie in zukünftigen Perioden als bekannt vorausgesetzt werden können. Ist eine Regel bereits bekannt, d.h., existiert sie in der Regelbasis, so muss ein Update dieser Regel bzgl. der Historie der statistischen Parameter (insbesondere Support und Confidence) erfolgen. Anhand dieser Historie können Trends abgeleitet und ggf. geeignete Maßnahmen initialisiert werden; ferner kann der Erfolg bestimmter Maßnahmen gemessen werden. Unter der Annahme, dass eine Warenkorbanalyse die Regel $A \Rightarrow B$ ergibt, wird z.B. eine Promotion-Aktion für A gestartet, die ihrerseits das Ziel verfolgt, B im Verlauf mitzuziehen. Ergibt sich nun aus der Historie dieser Regel zwar ein konstanter Support, dagegen im Verlauf der Aktion ein Sinken der Confidence, so wird das Nichterreichen des anvisierten Ziels deutlich. Vielmehr haben die loyalen Käufer weiterhin beide Produkte gemeinsam gekauft, die durch die Aktion angelockten Käufer hingegen haben sich lediglich für A entschieden (vgl. Agrawal u. a., 1993). Um solche Entwicklungen in der Regelhistorie aufzudecken, verwenden Agrawal und Psaila (1995) sogenannte *Trigger*, die auf Veränderungen in der Historie reagieren und die sich entwickelnden Regeln (*emerging patterns*) aufdecken (vgl. Dong und Li, 1999).

Definition 2.5. Jede Regel $r_{i'}^{t'}$, $t' > t$, heißt Regel zu einem *emerging pattern* bzgl. r_i^t, wenn sie die folgenden Bedingungen erfüllt:

- Bedingung und Folgerung sind in r_i^t und $r_{i'}^{t'}$ identisch, d.h. zu den den untersuchten Zeitpunkten t und t'.
- Die Supports der zwei Regeln unterscheiden sich signifikant.

Die Trigger vermögen jedoch nur, Veränderungen in den statistischen Parametern zu erfassen und daraus mögliche Trends abzuleiten sowie neue Regeln zu erkennen, die von den bisher bekannten abweichen. Vergleichbar beschreiben es Chakrabarti u. a. (1998): Diese Autoren entwickelten einen Ansatz zum Aufdecken sogenannter *surprising patterns*, d.h. Muster von noch nicht vorgekommener und unerwarteter Art. Dieses Vorgehen weist einen nahezu statischen Charakter auf, da lediglich unterschiedliche Gruppen untersucht und ihre Unterschiede analysiert werden. Ein solcher Ansatz zur getrennten Analyse ohne Apriori-Wissen wird auch von Bay und Pazzani (1999) geliefert.

Ein anderes Vorgehen stellen Raghavan und Hafez (2000) vor. Sie untersuchen, inwiefern Regeln in zwei aufeinanderfolgenden Zeitintervallen ausreichenden Support besitzen. Dabei

unterscheiden sie drei wesentliche Arten von Itemmengen: häufige, abnehmend-häufige und zunehmend-häufige Itemmengen. Unter dem Begriff *häufige Itemmengen* versteht man solche, die über *alle* betrachteten Zeitintervalle die in Definition 2.3 gegebene Voraussetzung erfüllen; abnehmend-häufige Itemmengen entsprechen Itemmengen, die in den vorangegangen Zeitintervallen häufig waren, deren Support jedoch in der aktuellen Periode in gewissem Maße abgenommen hat:

Definition 2.6. Sei A eine häufige Itemmenge (oder zumimndest zunehmend-häufige Itemmenge, vgl. Definition 2.7) in einer Transaktionsteilmenge $\mathfrak{T}_t$, $t \geq 1$. A heißt *abnehmend-häufige Itemmenge* in der Transaktionsmenge $\mathfrak{T}_{t'}$, falls

$$Support_{\min} > \frac{\sum_{\upsilon_t=t_1}^{t_2} |\mathfrak{T}_{\upsilon_t}| \, (Support_{\min} + \pi_{\upsilon_t})}{\sum_{\upsilon_t=t_1}^{t_2} |\mathfrak{T}_{\upsilon_t}|} \geq \frac{Support_{\min}}{\alpha_{AR}}$$

$\forall t < t_2 \leq t^{'}$, wobei
- $1 \leq t_1 \leq t_2$,
- $1 \leq \alpha^{AR} \leq \infty$,
- $Support(A) = Support_{\min} + \pi$ mit $\pi \geq 0$, falls A häufige Itemmenge ist, und andernfalls $\pi < 0$.

Analog dazu handelt es sich bei zunehmend-häufigen Itemmengen um Itemmengen, die im vorangegangen Zeitintervall nicht häufig waren, d.h. entweder selten oder aber abnehmend-häufig, im aktuellen jedoch gewissen Häufigkeitsbedingungen genügen.

Definition 2.7. A heißt *zunehmend-häufige Itemmenge* in Transaktionsteilmenge $\mathfrak{T}_t, t > 1$, falls A seltene Itemmenge in Transaktionteilmenge $\mathfrak{T}_{t-1}$ war und $|\mathfrak{T}_t| \, \pi_t \geq 0$ gilt oder falls A abnehmend-häufige Itemmenge in $\mathfrak{T}_{t-1}$ war, $t > 1$, und die Bedingung $\sum_{\upsilon_t=t_1}^{t_2} |\mathfrak{T}_{\upsilon_t}| \, \pi_{\upsilon_t} \geq 0, t_1 \geq 1$ erfüllt ist.

Der in Definition 2.6 eingeführte Parameter α_{AR} legt dabei die Menge der beibehaltenen Vergangenheitsinformationen fest:
- $\alpha_{AR} = 1$: Verbannung jeder abnehmend-häufige Itemmenge A aus der Menge der häufigen Itemmengen (seltenes Itemset ohne Historie).
- $\alpha_{AR} \to \infty$: Erhalt jeder abnehmend-häufige Itemmenge A für zukünftige Berechnungen (seltenes Itemset mit Historie).
- α_{AR} nahe 1: Stärkere Fokussierung auf zunehmend-häufige als auf abnehmend-häufige Itemsets
  $\leftrightarrow$ α_{AR} weit von 1: entgegengesetztes Verhalten.

Des Weiteren wird neben der Festlegung des Grads an Vergangenheitsinformationen noch ein weiterer Parameter bestimmt, die sogenannte *Lokalität*. Mit Hilfe dieses Parameters gelingt es, die Signifikanz und die Relevanz der generierten abnehmend- oder zunehmend-häufigen Itemmengen zu bestimmen sowie große Itemmengen mit geringeren Supportwerten zu erzeugen, ohne in jeder Periode eine vollständige Analyse durchzuführen (vgl. Raghavan und Hafez, 2000). Die Lokalität setzt dabei die Größe der Transaktionsteilmengen derjenigen Zeitintervalle, in denen A (zunehmend-)häufig war, in Verhältnis zu der Größe aller Transaktionsteilmengen (vgl. Definition 2.8) und überprüft so die Anzahl der Zeitintervalle, in denen eine Itemmenge als (zunehmend-)häufig angesehen werden kann, und damit ihre Signifikanz für den gesamten bisher betrachteten Zeitraum.

Definition 2.8. Für eine Itemmenge A und eine Transaktionsteilmenge $\mathfrak{T}_t$ ist die *lokalität* (A) definiert als das Verhältnis der totalen Größe der Transaktionsteilmengen, in denen A entweder häufig oder zunehmend-häufig ist, zur totalen Größe *aller* Transaktionsuntermengen *lokalität* $(A) = \mathfrak{T}_{\upsilon_t}, 1 < \upsilon_t \leq t$:

$$\frac{\sum_{\forall \upsilon_t : A \text{ ist häufige oder zunehmend-häufige Itemmenge}} |\mathfrak{T}_{\upsilon_t}|}{\sum_{\upsilon_t=1}^{t} |\mathfrak{T}_{\upsilon_t}|}$$

Die Lokalität nimmt ihren maximalen Wert *lokalität* $(A) = 1$ an, wenn A in allen vergangenen Zeitintervallen (zunehmend-)häufig war (vgl. Raghavan und Hafez, 2000).

Neben der Tatsache, dass Raghavan und Hafez (2000) die Berechnungen bezogen auf alle Vergangenheitsdaten durchführen, berücksichtigt dieser Ansatz ebenso wie die zuvor eingeführten nicht die Möglichkeit, dass sich Regeln sowohl bzgl. ihrer Bedingungen als auch ihrer Folgerungen verändern können. Diese Möglichkeit wird u.a. von Song u. a. (2001) und darauf aufbauend von Chen u. a. (2005) betrachtet, die auf Basis von Agrawal und Psaila (1995) die Analyse dynamischer Veränderungen von Assoziationsregeln erweitern. Beide richten ihr Augenmerk dabei auf den Zusammenhang zwischen Konsumenteneigenschaften und der daraus resultierenden Kaufentscheidung; die Ansätze zur Dynamisierung lassen sich jedoch o.B.d.A. auf weitere Anwendungsgebiete übertragen. Im Gegensatz zu Agrawal und Psaila (1995) analysieren Song u. a. (2001) und Chen u. a. (2005) neben den *emerging patterns* (Definition 2.5) weitere potentielle Veränderungen der Regelmenge R und können so nicht nur Trends auf Basis der statistischen Parameter, sondern auch strukturelle Unterschiede innerhalb der Regeln aufdecken, sogenannte *unerwartete Veränderungen* (vgl. Liu und Hsu, 1996). Eine Addition von Regeln zur Basis ist dabei ebenso möglich wie ihr Sterben bei mangelndem Support entsprechend dem Ansatz von Raghavan und Hafez (2000). Song u. a. (2001) und Chen u. a. (2005) unterscheiden sich darin, dass sich Song u. a. (2001) auf einattributige Folgerungen beschränken, wohingegen Chen u. a. (2005) die Möglichkeit einbeziehen, dass eine Folgerung aus mehreren Attributen besteht. Insgesamt lassen sich also die folgenden vier Arten von Veränderungen in der Regelbasis festhalten:

1. *emerging patterns*: Bedingung und Folgerung bleiben konstant, der Support des Musters verändert sich jedoch signifikant (Definition 2.5, (vgl. Agrawal und Psaila, 1995; Dong und Li, 1999))
2. *hinzugefügte Regeln*: Sowohl die Bedingung als auch die Folgerung der untersuchten Regeln unterscheiden sich signifikant von allen in der Regelbasis vorhandenen Regeln (Definition 2.9, (vgl. Song u. a., 2001)).
3. *zerstörte Regeln*: Keine der neugelernten Regeln stimmt in Bedingung und Folgerung mit einer in der Regelbasis vorhandenen Regel überein. Es handelt sich also um ein Muster, das in der Gegenwart nicht länger zu finden ist (Definition 2.9).
4. *unerwartete Veränderungen*: Vorhandene Regeln weisen innerhalb ihrer Bedingung bzw. Folgerung Veränderungen auf; diese Art der Veränderungen wird im Zusammenhang mit sogenannten *interessanten* Mustern untersucht (vgl. Liu und Hsu, 1996; Silberschatz und Tuzhilin, 1996).

Definition 2.9. Eine Regel $r_{i'}^{t'}$ heißt *hinzugefügte Regel*, falls Bedingung und Folgerung sich deutlich von allen Regeln r_i^t in R^t unterscheiden. r_i^t heißt *zerstört*, falls Bedingung und Folgerung sich deutlich von allen Regeln $r_{i'}^{t'}$ in $R^{t'}$ unterscheiden.

Bei den unerwarteten Veränderungen muss zwischen zwei möglichen Arten der Veränderungen unterschieden werden, d.h. wie weit sich die Regeln innerhalb der Bedingung oder aber der Folgerung verändert haben (Definition 2.10). Man spricht bei dieser Form der Veränderung von dem Unterschied zwischen dem alten, in der Regelbasis vorhandenen *Glauben* eines Nutzers und der neu gelernten Regel (vgl. Song u. a., 2001).

Definition 2.10. Eine Regel $r_{i'}^{t'}$ heißt *unerwartete Veränderung der Folgerung* bzgl. r_i^t, falls die Bedingungen der Regeln r_i^t und $r_{i'}^{t'}$ gleichartig sind, ihre Folgerungen sich jedoch unterscheiden (vgl. Liu und Hsu, 1996; Song u. a., 2001)). $r_{i'}^{t'}$ heißt *unerwartete Veränderung der Bedingung* bzgl. r_i^t, falls die Folgerungen der Regeln r_i^t und $r_{i'}^{t'}$ gleichartig sind, ihre Bedingungen sich jedoch unterscheiden (vgl. Chen u. a., 2005).

Um die möglichen Veränderungen zu analysieren, wird ein dreistufiges Vorgehen angewandt:

1. Association Rule Mining im klassischen (statischen) Sinn auf aktuellem Datensatz
2. Regel-Matching: Vergleich der bekannten mit den neu erlernten Regeln unter Hinzuziehung eines Ähnlichkeits- und eines Differenzmaßes zur Unterscheidung der einzelnen Veränderungsarten
3. Evaluierung des Grads der Veränderung

Der erste Schritt erfolgt mit Hilfe des aus dem statischen Association Rule Mining bekannten Apriori-Algorithmus. Hierbei ist zu beachten, dass für die dynamische Analyse ein geringer Minimalsupport sinnvoll ist, so dass emerging patterns eher erkannt werden können (vgl. Song u. a., 2001). Im zweiten Schritt werden die Veränderungen basierend auf ihrer maximalen Ähnlichkeit und dem Differenzmaß klassifiziert. Bei der Ähnlichkeitsbestimmung wird der Grad an Übereinstimmung zwischen zwei Regeln berechnet (vgl. Liu und Hsu, 1996; Song u. a., 2001); bei Einbeziehung von mehrattributigen Folgerungen erfolgt diese Berechnung mit Hilfe von (2.1) (vgl. Chen u. a., 2005).

$$s_{ii'}^{AR} = \begin{cases} \frac{\ell_{ii'} \sum_{q_1} y_{ii'q_1}}{|A_{ii'}|} \times \frac{\hbar_{ii'} \sum_{q_2} z_{ii'q_2}}{|B_{ii'}|} & , \text{ falls } |A_{ii'}| \neq 0 \wedge |B_{ii'}| \neq 0 \\ 0 & , \text{ sonst} \end{cases} , \tag{2.1}$$

wobei

- $A_{ii'}$ und $B_{ii'}$: Mengen an Attributen, die sowohl in r_i^t als auch in $r_{i'}^{t'}$ vorhanden sind; A – Bedingungen, B – Folgerungen
- $\ell_{ii'}$: Ähnlichkeit der Bedingungen; $\ell_{ii'} = \frac{|A_{ii'}|}{\max(|A_i^t|,|A_{i'}^{t'}|)}$, dabei ist $|A_\cdot|$ die Anzahl an Attributen der Bedingungen zum jeweiligen Zeitpunkt.
- $\hbar_{ii'}$: Ähnlichkeit der Folgerungen; $\hbar_{ii'} = \frac{|B_{ii'}|}{\max(|B_i^t|,|B_{i'}^{t'}|)}$, dabei ist $|B_\cdot|$ die Anzahl an Attributen der Folgerungen zum jeweiligen Zeitpunkt.
- $y_{ii'q_1}$: Binärvariable – 1, wenn q_1-tes Attribut in $A_{ii'}$ für r_i^t und $r_{i'}^{t'}$ denselben Wert annimmt, 0 sonst.
- $z_{ii'q_2}$: Binärvariable – 1, wenn q_2-tes Attribut in $B_{ii'}$ für r_i^t und $r_{i'}^{t'}$ denselben Wert annimmt, 0 sonst.

Für die Ähnlichkeit gilt $s_{ii'}^{AR} \in [0,1]$, wobei für identische Regeln $s_{ii'}^{AR} = 1$ gilt. Im Anschluss an die Bestimmung der Ähnlichkeiten zwischen den einzelnen Regeln in R^t und $R^{t'}$ werden die

maximalen Ähnlichkeitswerte der einzelnen Regeln ermittelt und mit einem vorgegeben Grenzwert RMT (*R*ule *M*atching *T*hreshold, (vgl. Song u. a., 2001; Chen u. a., 2005)) verglichen, um hinzugefügte und zerstörte Regeln zu identifizieren:

- $s_{i\cdot}^{AR} = \max\left(s_{i1}^{AR}, \ldots, s_{i|R^{t'}|}^{AR}\right)$: maximale Ähnlichkeit für r_i^t. Falls $s_{i\cdot}^{AR} < RMT$ gilt, ist die Regel zum Zeitpunkt t' nicht mehr vorhanden, d.h., es handelt sich um eine zerstörte Regel.
- $s_{\cdot i'}^{AR} = \max\left(s_{1i'}^{AR}, \ldots, s_{|R^t|i'}^{AR}\right)$: maximale Ähnlichkeit für $r_{i'}^{t'}$. Falls $s_{\cdot i'}^{AR} < RMT$ gilt, ist die Regel nicht in der Regelbasis R^t vorhanden, d.h., es handelt sich um eine hinzugefügte Regel.

Übersteigt der ermittelte maximale Ähnlichkeitswert jeweils den Grenzwert RMT, so muss die neue Regel einer in der Regelbasis vorhandenen Regel zugeordnet werden. Es wird dementsprechend jeweils überprüft, inwiefern es sich um dieselbe Regel mit unerwarteter Veränderung handelt. Hierzu wird das in (2.2) gegebene Differenzmaß verwendet.

$$\Delta\left(r_i^t, r_{i'}^{t'}\right) = \begin{cases} \frac{\ell_{ii'} \sum_{q_1} y_{ii'q_1}}{|A_{ii'}|} - \frac{\hbar_{ii'} \sum_{q_2} z_{ii'q_2}}{|B_{ii'}|} & , \text{ falls } |A_{ii'}| \neq 0 \wedge |B_{ii'}| \neq 0 \\ 0 & , \text{ sonst} \end{cases} \tag{2.2}$$

Aufgrund der Berechnung der Differenz zwischen der Ähnlichkeit der Bedingungen und der Ähnlichkeit der Folgerungen können die unterschiedlichen Arten unerwarteter Veränderungen unterschieden werden (vgl. Chen u. a., 2005; Song u. a., 2001):

- Ist $\Delta\left(r_i^t, r_{i'}^{t'}\right) > 0$, so enthält $r_{i'}^{t'}$ eine unerwartete Veränderung der Folgerung bzgl. r_i^t, d.h., $r_{i'}^{t'}$ enthält einen unerwarteten Kauf.
- Ist $\Delta\left(r_i^t, r_{i'}^{t'}\right) < 0$, so enthält $r_{i'}^{t'}$ eine unerwartete Veränderung der Bedingung bzgl. r_i^t, d.h., $r_{i'}^{t'}$ enthält eine unerwartete Käuferbewegung.
- Ist $\Delta\left(r_i^t, r_{i'}^{t'}\right) = 0$, so sind die Regeln r_i^t und $r_{i'}^{t'}$ entweder identisch ($\ell_{ii'} = 1$ und $\hbar_{ii'} = 1$) oder vollkommen unterschiedlich.

Der Fall $\Delta\left(r_i^t, r_{i'}^{t'}\right) = 0$ stellt kein Problem dar, da identische Regeln aufgrund ihres Ähnlichkeitswertes $s_{ii'}^{AR} = 1$ leicht identifiziert werden können. Problematisch wird es jedoch in dem Moment, in dem eine Regel $r_{i'}^{t'}$ bzgl. einer Regel $r_{i_1}^t$ ein emerging pattern, bzgl. einer anderen Regel $r_{i_2}^t$ jedoch unerwartet ist. In diesem Fall ist das emerging pattern und damit die Regel $r_{i_1}^t$ vorzuziehen, eine nicht direkt aus $\Delta\left(r_i^t, r_{i'}^{t'}\right)$ ablesbare Folgerung. Daher wird in Song u. a. (2001) eine Modifizierung des Differenzmaßes vorgenommen:

$$\Delta'\left(r_i^t, r_{i'}^{t'}\right) = \left|\Delta\left(r_i^t, r_{i'}^{t'}\right)\right| - \omega_{ii'}, \text{ wobei } \omega_{ii'} = \begin{cases} 1 & , \text{ falls } \max\left(s_{i\cdot}^{AR}, s_{\cdot i'}^{AR}\right) = 1 \\ 0 & , \text{ sonst} \end{cases}$$

Durch die Einbeziehung von $\omega_{ii'}$ können so bei der Bestimmung von potentiellen unerwarteten Veränderungen Regeln ausgeschlossen werden, die bereits als identisch zu einer Regel identifiziert wurden. Falls $\Delta'\left(r_i^t, r_{i'}^{t'}\right) \geq RMT$ gilt, enthält $r_{i'}^{t'}$ eine unerwartete Veränderung im Vergleich zu r_i^t[7]. Damit lassen sich die unterschiedlichen Veränderungen wie folgt identifizieren:

- *emerging pattern*: $s_{ii'}^{AR} = 1$

[7]Für unerwartete Veränderungen wird i.d.R. $RMT = 1$ gesetzt (vgl. Chen u. a., 2005).

- *hinzugefügte Regel*: $s_{\cdot i'}^{AR^{t+1}} < RMT$
- *zerstörte Regel*: $s_{i\cdot}^{AR^{t}} < RMT$
- *unerwartete Veränderung der Bedingung*: $\Delta\left(r_i^t, r_{i'}^{t'}\right) = -1, \Delta'\left(r_i^t, r_{i'}^{t'}\right) = 1$
- *unerwartete Veränderung der Folgerung*: $\Delta\left(r_i^t, r_{i'}^{t'}\right) = 1, \Delta'\left(r_i^t, r_{i'}^{t'}\right) = 1$

In einem letzten Schritt werden die Veränderungen bzgl. ihrer Signifikanz evaluiert, indem die Differenzen des Supports zu den einzelnen Zeitpunkten betrachtet werden. Im Falle eines emerging patterns wird die Wachstums- bzw. Niedergangsrate des Supports untersucht; bei hinzugefügten Regeln werden die jeweils maximalen Ähnlichkeitswerte $s_{i'}^{AR}$ bzw. s_i^{AR} einbezogen. Handelt es sich um eine unerwartete Veränderung, so werden die Vereinigungsmengen der jeweiligen Itemmengen in der Bedingung und der Folgerung betrachtet:

$$\chi_{ii'} = \begin{cases} \frac{Support^{t'}(r_i) - Support^{t}(r_i)}{Support^{t}(r_i)} & \text{für emerging patterns} \\ \frac{Support^{t'}(r_{i \cap i'})}{Support^{t'}(r_{i'})} & \text{für unerwartete Veränderungen} \\ \left(1 - s_i^{AR}\right) \times Support^{t}\left(r_i\right) & \text{für zerstörte Regeln} \\ \left(1 - s_{i'}^{AR}\right) \times Support^{t'}\left(r_{i'}\right) & \text{für hinzugefügte Regeln} \end{cases}$$

Im Bereich der Marktforschung können die ermittelten Veränderungen herangezogen werden, um Marketingressourcen gezielt zu verteilen; Chen u. a. (2005) stellen hierzu verschiedene Möglichkeiten vor. So können bei Mustern, die sich positiv entwickeln, d.h. bei emerging patterns mit wachsendem Support, gezielte Marketingmaßnahmen die Entwicklung fördern. Dagegen implizieren sich negativ entwickelnde oder gar zerstörte Regeln, dass die aufgrund der ermittelten Zusammenhänge eingesetzten Ressourcen anderweitig verwendet werden können; so können z.B. neu auftretende Kaufverhaltensmuster, d.h. hinzugefügte Regeln, unterstützt werden. Das Verhalten bei einer unerwarteten Veränderung hängt davon ab, auf welcher Seite der Regel die Veränderung auftritt. Sollte die Bedingung sich geändert haben, bedeutet dies, dass sich bei Analyse der Zusammenhänge zwischen Kundeneigenschaften und Kaufmustern der Kundenstamm, der die Regel repräsentiert, verändert hat. Unter Einbeziehung der Interpretationsmöglichkeiten eines einfachen Recommenders ist es möglich, dass im Laufe der Zeit ein Produkt in der Bedingung durch ein anderes, neueres Produkt substituiert wird. Bei einer Veränderung der Folgerung hingegen bedeutet es, dass die resultierenden Kaufmuster sich über die Zeit hinweg verändert haben. Als Konsequenz müsste dann untersucht werden, ob alte, aus der Regel weggefallene Produkte durch gezielte Maßnahmen gefördert werden müssen, oder aber, ob neue Produkte hinzugefügt wurden, die beispielsweise von den bereits durchgeführten Maßnahmen profitieren und keine gesonderten Promotion-Aktionen benötigen.

Ähnlichkeit ist nichts anderes als Wiederkehr.

E. Jünger

3

Einführung in das Fuzzy-Clustering

Beim Fuzzy-Clustering handelt es sich um ein weiteres Data Mining-Verfahren aus dem Bereich der Clusteranalyse, das im Rahmen des KDD-Prozesses zur Anwendung gelangt (vgl. Abschnitt 1.1). Unter der Bezeichnung *Clusteranalyse* wird eine Sammlung von Verfahren beschrieben, deren Zielsetzung in der Aufdeckung ähnlicher Eigenschaftsausprägungen von Objekten in vorhandenen Datensätzen besteht und die als Folge die Zusammenfassung von Objekten entsprechend ihrer Ähnlichkeit in Gruppen zulässt, den Clustern (vgl. Abschnitt 1.2). Zahlreiche Untersuchungen befassen sich mit der Clusteranalyse, nicht zuletzt wegen ihrer Vielzahl von Anwendungsmöglichkeiten in verschiedenen Bereichen (vgl. Aggarwal u. a., 2003; Zhou u. a., 2008). Dieses Kapitel stellt die relevanten Gebiete der Clusteranalyse vor, auf denen aufbauend eine dynamische Analyse der Clusterstruktur erfolgen kann.

3.1. Einführung in die Clusteranalyse

Als Ausgangspunkt einer Clusteranalyse dienen durch ihre repräsentativen Eigenschaften beschriebene Objekte; dabei können die einzelnen Attribute numerischer oder kategorischer Art sein. Handelt es sich um rein quantitative Daten, werden die Objekte durch ihre Eigenschaftsvektoren im p-dimensionalen Raum repräsentiert; ein Cluster entspricht dann demzufolge einer Anhäufung von Punkten in diesem Vektorraum. Um die Vergleichbarkeit zu gewährleisten, ist ein sinnvolles Skalierungsverhältnis der einzelnen Eigenschaftswerte unerlässlich, da unterschiedliche Skalen das Ergebnis verzerren oder ein gänzlich falsches erzeugen können. Im Idealfall kann eine Clustereinteilung intuitiv erfolgen. Abbildung 3.1 zeigt drei gut separierte Datengruppen und damit eine offensichtliche Clusterstruktur. Tatsächlich stellt sich die Struktur von Daten in der Realität meist komplexer dar: Zum einen handelt es sich i.d.R. nicht um zwei- bis drei-, sondern mehrdimensionale Objekte, d.h., die Objekte werden durch mehr als drei Eigenschaften repräsentiert, so dass eine grafische Darstellung nicht möglich ist, zum anderen gelingt auch die Separierung nicht immer so eindeutig wie in Abbildung 3.1. Zur Aufdeckung einer Clusterstruktur in komplexen Daten werden Clusteranalyseverfahren benötigt, die eine automatisierte Einteilung der Daten anhand ihrer Eigenschaften vornehmen (vgl. Bacher u. a., 2010, S. 18ff.). Dabei finden folgende Kriterien Beachtung:

- *Homogenität*: Die Ähnlichkeit innerhalb der Cluster soll maximal werden, d.h., die Distanzen zwischen den dem Cluster zugeordneten Objekten sollen möglichst gering sein.

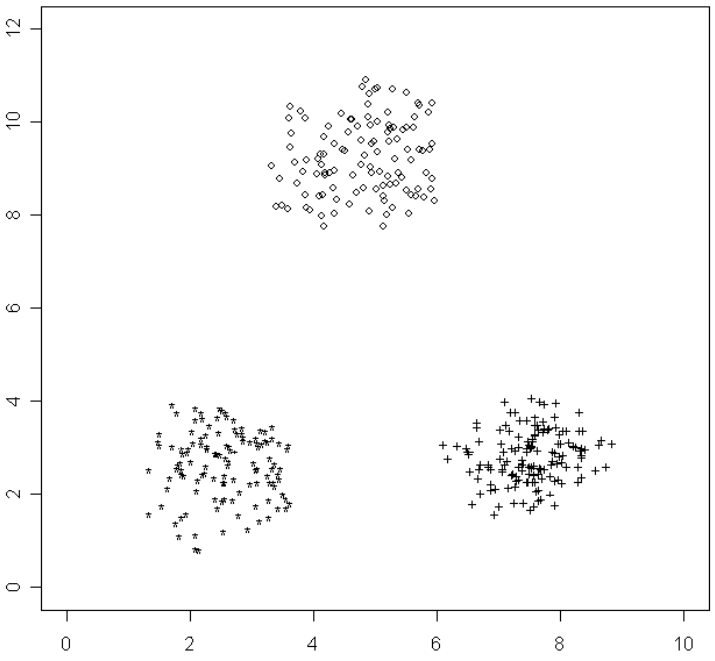


Abbildung 3.1.: Clusterstruktur im $\mathbb{R}^2$

- *Heterogenität*: Die Trennung der Cluster soll maximal sein, d.h., die Verschiedenheit der Cluster untereinander soll maximiert werden.

Mit Hilfe dieser beiden Kriterien soll die Anwendung einer möglichst eindeutigen Clustereinteilung gewährleistet werden.

Clusteranalyseverfahren, die auf diese Weise eine Strukturierung von Daten vornehmen, können in einer Vielzahl von Gebieten eingesetzt werden, unter anderem in folgenden:

- Im Marketing können Kunden- und Produktsegmente ermittelt werden, so dass bei Kundensegmenten mit ähnlichen Kaufverhaltensmustern z.B. die Möglichkeit besteht, die Marketingaktivitäten gezielt auf die ermittelten Segmente abzustimmen (vgl. u.a. Baier und Brusch, 2008, S. 771; Berry und Linoff, 2000, S. 13f.).
- Durch das Clustering von Clickstreams im Internet kann das Nutzerverhalten im Internet basierend auf der Navigationshistorie einzelner Besucher einer Seite analysiert werden, so dass sich die Möglichkeit bietet, die Navigation auf der Seite zu vereinfachen (vgl. u.a. Ali und Ketchpel, 2003; Banerjee und Ghosh, 2001).
- In der Medizin können Patienten mit ähnlichen Eigenschaften und Symptomen gruppiert werden, so dass entsprechend der Diagnose sinnvolle Behandlungsmaßnahmen bestimmt und eingeleitet werden können (vgl. Bramer, 2007, S. 221; Deichsel und Trampisch, 1985, S. 17).
- Im Rahmen des Umwelt-Monitorings können in Luftdaten Häufungspunkte besonderer Verschmutzungsgrade gemessen werden, so dass ggf. erforderliche Maßnahmen eingeleitet werden können (vgl. Minke und Lessing, 2010).

Die Wahl des zu verwendenden Verfahrens wird abhängig vom vorliegenden Datenmaterial und dem Ziel der Clusterzuordnung vorgenommen; man unterscheidet dabei zwischen hierarchischen und partitionierenden Verfahren. Bei den hierarchischen Verfahren erfolgt eine sukzessive

Entwicklung von Clustern, die auf zwei Arten vorgenommen werden kann: Zum einen kann eine Clusterhierarchie bestimmt werden, indem die gesamte Objektmenge nach und nach in kleinere Cluster zerlegt wird, bis im Extremfall jedes Objekt ein eigenes Cluster darstellt (divisives Vorgehen); zum anderen kann von Einobjekt-Clustern ausgegangen werden, die sukzessiv zusammengefasst werden, bis alle Objekte einem einzigen Cluster zugeordnet sind (agglomeratives Vorgehen). Das Ergebnis der hierarchischen Clusterverfahren lässt sich mit Hilfe eines Dendrogramms darstellen, aus dem die Clusterhierarchie ersichtlich wird (vgl. u.a. Jensen, 2008, S. 340f.).

Die partitionierenden Verfahren beginnen mit einer initialen Zerlegung der Objekte, von der ausgehend die Clustereinteilung verbessert werden soll; dabei ist die Clusterzahl typischerweise vorher definiert. Die Startzerlegung kann beliebig gewählt, kann jedoch auch das Ergebnis einer vorangegangen hierarchischen Analyse sein. Ziel der Analyse ist die Repräsentation eines Clusters durch einen Clusterprototypen, der die dem Cluster zugeordneten Objekte repräsentiert. Im einfachsten Fall handelt es sich bei dem Clusterprototypen um das Clusterzentrum, aber auch andere Clustereigenschaften wie Größe und Ausrichtung können darin enthalten sein. Diese Art von Verfahren basiert in der Regel auf einer Zielfunktion, die minimiert werden soll (vgl. Cios u. a., 2007, S. 258).

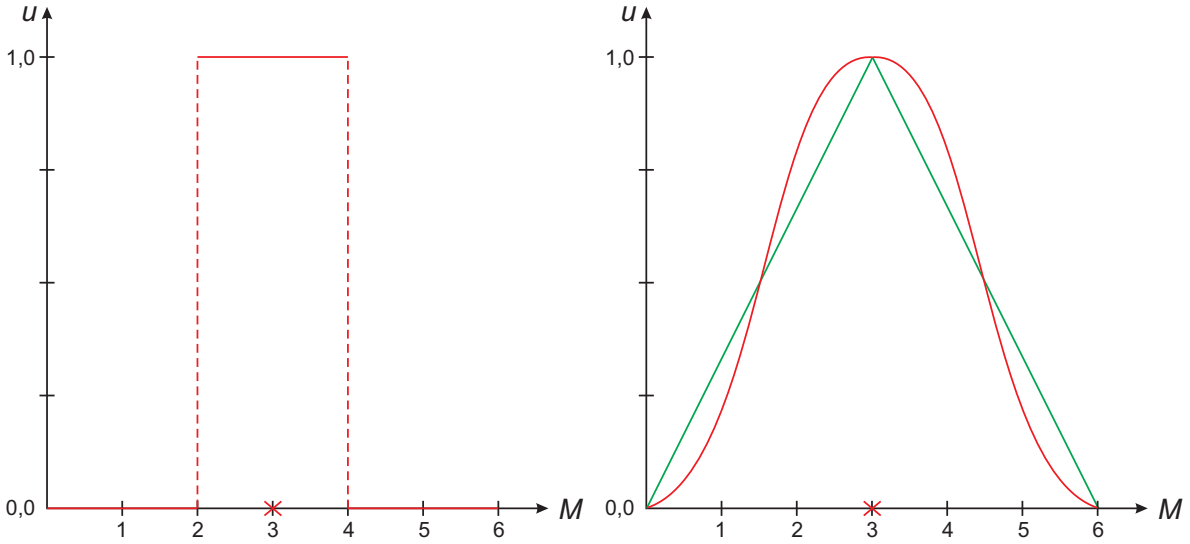
Clusterverfahren werden ferner in disjunkte und nichtdisjunkte Verfahren unterteilt (vgl. Bacher u. a., 2010, S. 147). Bei den disjunkten Verfahren schließen sich die Clusterzuordnungen aus, d.h., jedes Objekt wird genau einem Cluster zugeordnet; bei den nichtdisjunkten Verfahren ist hingegen eine Zuordnung zu mehreren Clustern möglich. Diese Zuordnung kann deterministischer Art sein oder anhand von Zugehörigkeitsgraden erfolgen. Im Falle der deterministischen Zuordnung ist ein Objekt entweder jeweils vollständig einem oder mehreren Clustern zugeordnet oder keinem. Bei der Zuordnung unter Einbeziehung von Zugehörigkeitsgraden $\in [0, 1]$ geben diese an, wie weit die Zugehörigkeit eines Objektes zu einem Cluster wahrscheinlich bzw. möglich ist. Das bekannteste disjunkte Verfahren ist das k-Means-Verfahren (vgl. u.a. Bacher u. a., 2010, S. 299ff.; Bramer, 2007, S. 224 ff.), das im Anschluss noch Erwähnung finden wird. Zu den nichtdisjunkten Verfahren zählen neben den modellbasierten Verfahren (Model Based Clustering), die von verschiedenen Verteilungsmodellen (sogenannten *Mixture Models*) ausgehend eine Clusterstruktur ermitteln (vgl. z.B. Cios u. a., 2007, S. 269f.), auch die Verfahren der Fuzzy-Clusteranalyse, die in Abschnitt 3.3 näher betrachtet werden.

3.2. Fuzzy-Logik

Als Grundlage für das Fuzzy-Clustering dient neben der Clusteranalyse der Bereich der Fuzzy-Logik[8]; dieser Begriff wurde erstmals von Zadeh (1965) eingeführt. Im Gegensatz zur klassischen Aussagenlogik lässt die Fuzzy-Logik neben den harten Aussagen *wahr* und *falsch* bzw. Null und Eins unscharfe Abstufungen zu und ermöglicht es damit, semantische Ausdrücke wie beispielsweise *weitestgehend*, *viel* oder *wenig* mathematisch zu modellieren (vgl. Deimer, 1986, S. 115f.). Die Notwendigkeit einer solchen mathematischen Modellierung soll anhand eines vereinfachten, eindimensionalen Beispiels erläutert werden: Angenommen, ein Produktsegment sei durch sein

[8] *fuzzy* (engl.): undeutlich, unscharf, verschwommen

Idealprodukt bzgl. eines Merkmals M mit der Merkmalsausprägung $M = 3$ repräsentiert. Mit Hilfe der klassischen Aussagenlogik müssen Grenzwerte bestimmt werden, so dass alle Realprodukte, die eine ausreichende Ähnlichkeit vorweisen, diesem Produktsegment zugeordnet werden (z.B. $2 \leq M \leq 4$). Als Ergebnis ergibt sich für jedes Realprodukt eine eindeutig vorhandene oder eindeutig fehlende Zuordnung (vgl. Abbildung 3.2a). Da kaum eine Begründung dafür gegeben werden kann, dass ein Produkt mit einem Wert $M = 2$ gleichermaßen diesem Produktsegment zuzuordnen ist wie ein zum Idealprodukt identisches, während ein Produkt mit $M = 1,99$ als gänzlich verschieden zu diesem Segment angenommen wird, scheint diese Zuordnung wegen ihrer harten Grenzen eher willkürlich gewählt. Werden dagegen anstelle der harten Mengen Fuzzy-Mengen verwendet (vgl. Definition 3.1), gelingt es, durch Ermittlung der Zugehörigkeitsgrade einen fließenden Übergang zu erzeugen. Dabei sind verschiedene Möglichkeiten denkbar, die Zugehörigkeitsgrade mit Hilfe einer Fuzzy-Menge zu modellieren (vgl. Abbildung 3.2b).



(a) Harte Zuordnung (b) Unscharfe Zuordnung

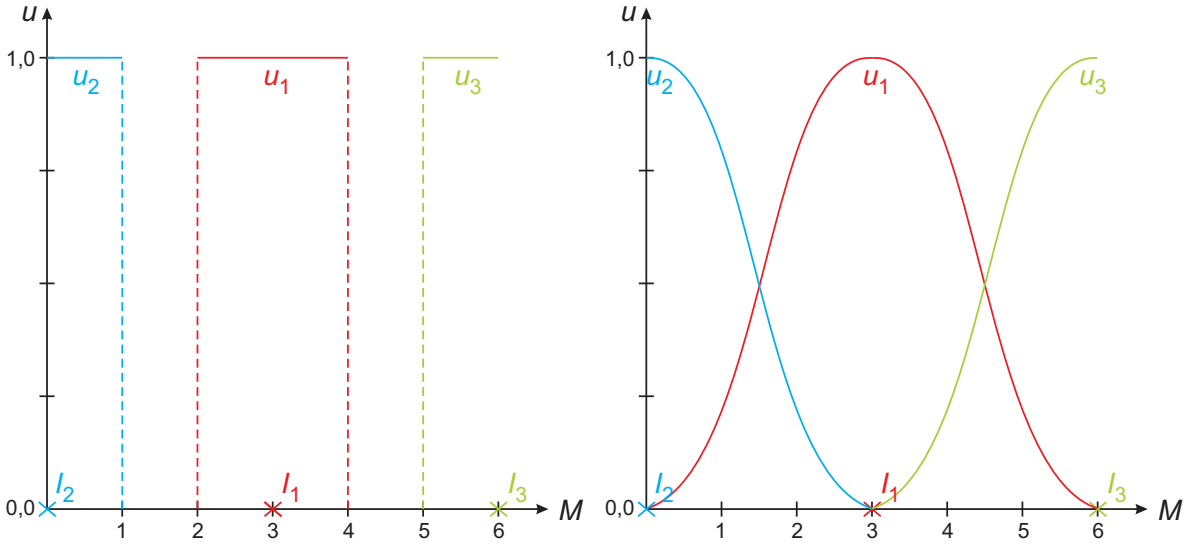
Abbildung 3.2.: Harte und unscharfe Zuordnung zu einem Produktsegment

Definition 3.1. (Kruse u. a., 1993, S. 10) Eine Fuzzy-Menge u von O ist eine Funktion von der Grundmenge O in das Einheitsintervall, d.h.

$$u : O \rightarrow [0, 1].$$

$F(O)$ bezeichnet die Menge aller Fuzzy-Mengen von O.

Auf diese Weise gelingt es, die Zugehörigkeit zu verschiedenen Produktsegmenten fließend zu modellieren. In Abbildung 3.3 werden drei Produktsegmente durch ihr jeweiliges Idealprodukt I repräsentiert, so dass im Falle der unscharfen Zuordnung ein gradueller Übergang zwischen den Segmenten stattfindet; ein Produkt, das sich direkt zwischen zwei Idealprodukten befindet, erhält hier denselben Zugehörigkeitsgrad zu beiden Segmenten (vgl. Abbildung 3.3b). Der Zugehörigkeitsgrad drückt aus, wie stark ein Objekt einem Cluster zuzuordnen ist; ein Zugehörigkeitsgrad von Null entspricht analog zur harten Aussagenlogik einer nicht vorhandenen



(a) Harte Zuordnung (b) Unscharfe Zuordnung

Abbildung 3.3.: Harte und unscharfe Zuordnung bei drei Produktsegmenten

Zugehörigkeit, während ein Zugehörigkeitsgrad von Eins eine totale Zugehörigkeit darstellt. Dabei werden zwei Arten von Zugehörigkeitsgraden unterschieden: Probabilistische Zugehörigkeitsgrade geben die Wahrscheinlichkeit an, mit der ein Objekt zu einem Cluster gehört; die einzelnen Zugehörigkeitsgrade lassen sich demnach zu einem Gesamtwert von Eins aufaddieren. Possibilistische Zugehörigkeitsgrade hingegen sind als Möglichkeit einer Zuordnung zu interpretieren, so dass eine Mehrfachzuordnung zu verschiedenen Clustern mit hohen Zugehörigkeitsgraden oder aber eine fehlende Zuordnung zu irgendeinem Cluster auftreten kann (vgl. Zadeh, 1978). Die Fuzzy-Logik hat im Laufe der Zeit in verschiedene Bereiche wie z.B. in die Regelungstechnik, in die Bildverarbeitung (vgl. u.a. Kruse u. a., 1993, S. 69) sowie in die Datenanalyse Einzug erhalten.

3.3. Probabilistische Clusteranalyse

Bei der Fuzzy-Clusteranalyse oder kurz dem Fuzzy-Clustering wird der Ansatz der Clusteranalyse mit der Grundidee der Fuzzy-Logik verbunden; dieser Ansatz geht zurück auf Dunn (1973). Auf eindeutige, harte Zuteilungen zu den Clustern wird verzichtet, so dass graduelle Zugehörigkeiten zu den verschiedenen Clustern ermöglicht werden. Dies ist z.B. im Rahmen einer Kundensegmentierung sinnvoll, da ein einzelner Kunde typische Charakteristika verschiedener Segmente in sich vereinen kann. Abbildung 3.4 soll diesen Umstand verdeutlichen: Intuitiv sind zwei Cluster zu erkennen, die die gleiche Struktur besitzen. Als problematisch erweist sich jedoch die Zuordnung des Objektes, das sich in der Schnittmenge beider Clustern befindet. Erfolgte die Zuweisung zu einem der beiden Cluster, ließe sich anhand der harten Zuordnungen die symmetrische Struktur der Daten nicht länger erkennen. Durch die Verwendung von Zugehörigkeitsgraden kann dieses Objekt jedoch gleichermaßen beiden Clustern zugeordnet werden, so dass ein detaillierteres Muster entsteht und die grundsätzliche Information bzgl. der Symmetrie der Datenstruktur erhalten bleibt (vgl. Kruse u. a., 2007, S. 14).

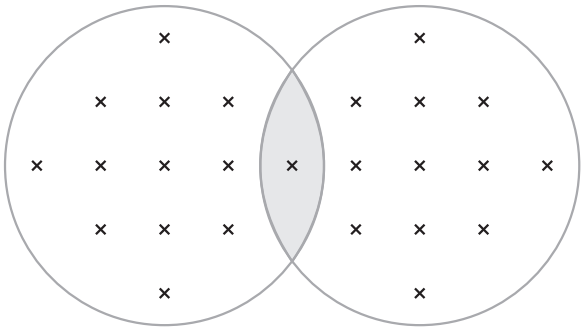


Abbildung 3.4.: Clusterstruktur bei zwei Clustern

Das Fuzzy-Clustering zählt zu den partitionierenden Verfahren (vgl. Abschnitt 3.1). Ziel ist die Bestimmung von Clusterprototypen C_i[9], die die einzelnen Cluster repräsentieren, indem eine Zielfunktion unter Berücksichtigung bestimmter Restriktionen optimiert wird. Für alle partitionierenden Clusterverfahren gilt die Bedingung, dass die einem Clusterprototypen zugeordneten Objekte aufgrund der Homogenitätsbedingung diesem möglichst ähnlich sein sollen. Die Eigenheit des Fuzzy-Clusterings besteht jedoch darin, dass nicht alle Distanzwerte gleich gewichtet sind; vielmehr erfolgt eine Gewichtung der Distanzen der Objekte zum Clusterprototypen über den jeweiligen Zugehörigkeitsgrad. Begründet wird diese Tatsache dadurch, dass ein Objekt, das nur über eine geringe Zugehörigkeit verfügt, einen entsprechend niedrigen Einfluss auf die Positionierung des Clusterprototypen ausüben soll, während ein Objekt mit hoher Zugehörigkeit diese stark beeinflussen soll; als Folge ergibt sich die in (3.1) gegebene Zielfunktion.

$$J\left(X,U,C\right)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_j\right), \tag{3.1}$$

wobei

- X= (x_{jp}): $n \times p$-Matrix der Objekte bei n Obejtekn und p Dimensionen
- C= $\{C_1,\ldots,C_c\}$: Menge der Clusterprototypen bei c Clustern
- U= (u_{ij}): $c \times n$-Matrix der Zugehörigkeitsgrade
- $m\in(1,\infty)$: Fuzzifier zur Steuerung der Unschärfe[10]; i.d.R. $m=2$
- $d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)$: quadrierte Distanz zwischen Clusterzentrum $\vec{v}_i$ und Objekt $\vec{x}_j$ unter Berücksichtigung der Eigenschaften des Clusterprototypen C_i

Da es sich bei der gegebenen und den folgenden Zielfunktionen um eine Form der Heterogenitätsbewertung innerhalb einzelner Cluster anhand der gewichteten Distanzen zum Clusterzentrum handelt, sind diese im Laufe des Verfahrens zu minimieren.

Bei der ursprünglichen Form des Fuzzy-Clusterings wird von einem probabilistischen Ansatz ausgegangen, d.h., die Analyse basiert auf der Prämisse, dass eine Wahrscheinlichkeitsverteilung der Daten über die Cluster vorliegt (vgl. Höppner u. a., 1999, S. 28ff.). Die Zugehörigkeitsgrade

[9] C_i ist ein Platzhalter, der die wesentlichen Eigenschaften des i-ten Clusters enthält. Diese sind vom jeweiligen Verfahren abhängig.

[10] Je größer m gewählt wird, desto unschärfer ist die Partitionierung; je kleiner m, desto eher entspricht das Ergebnis einer harten Einteilung. Nähere Informationen zum Fuzzifier sind u.a. Klawonn und Höppner (2003) zu entnehmen.

werden daher im Sinne der Probabilitätstheorie als Wahrscheinlichkeiten oder Wahrheitsgrad einer Zuordnung interpretiert. Um diesem Umstand gerecht zu werden, müssen bei Optimierung der Zielfunktion in (3.1) die folgenden Restriktionen berücksichtigt werden:

$$\sum_{j=1}^{n} u_{ij} > 0 \; \forall \; i = 1 \ldots, c \tag{3.2}$$

$$\sum_{i=1}^{c} u_{ij} = 1 \; \forall \; j = 1, \ldots, n \tag{3.3}$$

Die erste Bedingung in (3.2) fordert, dass keines der resultierenden Cluster leer ist, sondern jedem Cluster Objekte zugeordnet werden. Die zweite Bedingung erzwingt eine Normierung der Zugehörigkeitsgrade im probabilistischen Kontext, so dass die Zugehörigkeitsgrade als Wahrscheinlichkeiten interpretiert werden können. Dabei bleibt anzumerken, dass die Wahrscheinlichkeit, mit der ein Objekt einem Cluster zugeordnet wird, nichts über die Qualität der Repräsentation durch diesen Clusterprototypen aussagt (vgl. Höppner u. a., 1999, S.19 sowie Abschnitt 3.4).

Als bekanntester Algorithmus firmiert der Fuzzy-C-Means nach Dunn (1973), der später von Bezdek (1980) verallgemeinert wurde. Der Fuzzy-C-Means baut auf dem aus der harten Analyse bekannten k-Means auf (vgl. Abschnitt 3.1) und wird häufig auch als dessen *Fuzzy*-Variante bezeichnet. Gemäß dem k-Means wird beim Fuzzy-C-Means die quadrierte euklidische Distanz verwendet, um die Verschiedenheit der Objekte zum jeweiligen Clusterprototypen zu bestimmen; die Repräsentation des Clusterprototypen C_i erfolgt dabei über das Clusterzentrum $\vec{v}_i$, d.h. $C_i = \vec{v}_i$. Die Optimierung der Zielfunktion in (3.1) hinsichtlich der Clusterzentren $\vec{v}_i$ sowie der Zugehörigkeitsgrade u_{ij} erfordert ein iteratives Vorgehen, das in Algorithmus 3.1 vorgestellt wird. Die in (3.4) und (3.5) angeführten Update-Regeln ergeben sich dabei aus der Minimierung der Zielfunktion (vgl. z.B. Cios u. a., 2007, S. 268f.; Höppner u. a., 1999, S. 20ff.).

Die Komplexität des probabilistischen Fuzzy-C-Means liegt bei $\mathrm{O}\left(n \cdot c\right)$ (vgl. u.a. Timm, 2002, S. 18). Trotz des iterativen Vorgehens und einer mangelnden Garantie für eine Konvergenz gegen das globale Minimum handelt es sich um ein robustes Vorgehen. Die Wahrscheinlichkeit für das Auffinden eines globalen Minimums kann deutlich erhöht werden, indem die Initialisierung mit unterschiedlichen Startverteilungen erfolgt. Als problematisch erweist sich am gegebenen Algorithmus neben der Annahme ähnlicher Clustergrößen, dass durch die Verwendung der quadrierten euklidischen Distanz eine weitestgehend kreis- oder kugelförmige Form aller Cluster impliziert wird. Dies trifft jedoch nicht in jedem Fall zu; gerade im Bereich der Marktsegmentierung stellen z.B. ellipsoide Cluster keine Seltenheit dar. Aufbauend auf dem Fuzzy-C-Means entwickelten daher Gustafson und Kessel (1979) einen Algorithmus, der die Form und die Ausrichtung der Cluster durch Anwendung eines clusterspezifischen Distanzmaßes einbezieht und damit die Berücksichtigung der Datenstruktur verbessert. Als Distanzmaß wird die Mahalanobis-Distanz verwendet (vgl. Mahalanobis (1930) bzw. Mahalanobis (1936)):

$$d^2_{C_i}\left(\vec{v}_i, \vec{x}_j\right) = \left(\vec{x}_j - \vec{v}_i\right)^{\mathrm{T}} \Sigma_i^{-1} \left(\vec{x}_j - \vec{v}_i\right), \tag{3.6}$$

wobei Σ_i die Kovarianzmatrix von Cluster i darstellt.

Der Clusterprototyp wird durch das Zentrum und die Kovarianzmatrix definiert, d.h. $C_i = (\vec{v}_i, \Sigma_i)$. Dementsprechend handelt es sich beim Fuzzy-C-Means um einen Spezialfall des Gustaf-

Algorithmus 3.1 Probabilistischer Fuzzy-C-Means (*probFCM*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix
- Initialisiere probabilistische Verteilung $U^{(0)}$

repeat {für $\rho = 1, \ldots$}

Update der Clusterzentren $C_i^{(\rho)} = \vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho-1)}$ mit

$$\vec{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \vec{x}_j}{\sum_{j=1}^{n} u_{ij}^m} \tag{3.4}$$

Update der Distanzmatrix $D^{(\rho)} = \left(d^2_{C_i^{(\rho)}} (\vec{v}_i, \vec{x}_j) \right)$ bei konstanten $C_i^{(\rho)}$ mittels quadrierter euklidischer Distanz

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ mit

$$u_{ij} = \frac{1}{\sum_{i'=1}^{c} \left(\frac{d^2_{C_i^{(\rho)}} (\vec{v}_i, \vec{x}_j)}{d^2_{C_i^{(\rho)}} (\vec{v}_{i'}, \vec{x}_j)} \right)^{\frac{1}{m-1}}} \tag{3.5}$$

until $\left\| U^{(\rho-1)} - U^{(\rho)} \right\| < \epsilon$

son-Kessel-Algorithmus, bei dem $\Sigma_i = E \; \forall i = 1, \ldots, c$ gilt, wobei E die Einheitsmatrix repräsentiert.

Die Größe der einzelnen Cluster ist vorzugeben, da bei beliebig „groß“ gewählten Σ_i die Gefahr besteht, dass alle Distanzen vernachlässigbar gering werden. Mangels Vorwissen zur tatsächlichen Clustergröße wird die Kovarianzmatrix zur Berechnung der Distanzen häufig normiert (vgl. Höppner u. a., 1999, S. 43). Die Distanzfunktion ergibt sich damit zu

$$d^2_{C_i} (\vec{v}_i, \vec{x}_j) = \det (\Sigma_i)^{\frac{1}{p}} (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} \Sigma_i^{-1} (\vec{x}_j - \vec{v}_i), \tag{3.7}$$

wobei p die Anzahl der Dimensionen angibt.[11] Aus der Verwendung von (3.7) als Distanzfunktion ergibt sich Algorithmus 3.2. Die Initialisierung erfolgt dabei häufig auf Basis einiger Iterationen des probabilistischen Fuzzy-C-Means (Algorithmus 3.1), da der Gustafson-Kessel-Algorithmus durch die Berechnung der Inverse der Kovarianzmatrix im Vergleich wesentlich aufwändiger ist.

Um neben der ellipsoiden Form der Cluster auch die Größe und Dichte einbeziehen zu können, haben Gath und Geva (1989) eine Erweiterung des Gustafson-Kessel-Algorithmus eingeführt.

[11] Durch Substitution der Kovarianzmatrizen mit der positiv definiten Normmatrix $A_i = \det (\Sigma_i)^{\frac{1}{p}} \Sigma_i^{-1}$ lässt sich die Distanzfunktion (3.7) vereinfachen zu $d^2_{C_i} (\vec{v}_i, \vec{x}_j) = (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i)$ (vgl. Timm, 2002, S. 20).

Algorithmus 3.2 Probabilistischer Gustafson-Kessel-Algorithmus (*probGK*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix
- Initialisiere auf Basis des probabilistischen Fuzzy-C-Means:
  - Probabilistische Verteilung $U^{(0)} = U^{probFCM}$ und
  - Clusterprotypen $C_i^{(0)} = C_i^{probFCM}$, $i = 1, \ldots, c$

repeat {für $\rho = 1, \ldots$}

Update der Kovarianzmatrizen $\Sigma_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho-1)}$ und konstanten $\vec{v}_i$ gemäß $C_i^{(\rho-1)}$

$$\Sigma_i = \frac{\sum_{j=1}^{n} u_{ij}^m \left(\vec{x}_j - \vec{v}_i\right) \left(\vec{x}_j - \vec{v}_i\right)^{\mathrm{T}}}{\sum_{j=1}^{n} u_{ij}^m} \tag{3.8}$$

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho-1)}$ nach (3.4)

Setze $C_i^{(\rho)} = \left(\vec{v}_i, \Sigma_i\right), i = 1, \ldots, c$

Update der Distanzen $D^{(\rho)} = \left(d^2_{C_i^{(\rho)}}\left(\vec{v}_i, \vec{x}_j\right)\right)$ bei konstanten $C_i^{(\rho)}$ nach (3.7)

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ nach (3.5)

until $\left\|U^{(\rho-1)} - U^{(\rho)}\right\| < \epsilon$

Sie gehen davon aus, dass alle Cluster einer eigenen Normalverteilung $N_i(\vec{v}_i, \Sigma_i)$ unterliegen, weiterhin werden die Clusterprototypen neben Clusterzentrum und Kovarianzmatrix durch ihre Apriori-Wahrscheinlichkeit P_i repräsentiert, d.h. $C_i = (\vec{v}_i, \Sigma_i, P_i)$. Die Interpretation der Daten erfolgt dabei als Realisierungen der p-dimensionalen normalverteilten Zufallsvariablen; die Distanzen werden umgekehrt proportional zum Maximum Likelihood-Schätzer der Aposteriori-Wahrscheinlichkeit bestimmt. Aufgrund der Verwendung des Maximum Likelihood-Schätzers zur Distanzbestimmung nähern sich Zugehörigkeitsgrade dabei einer harten Zuordnung an, d.h., sie tendieren gegen Null bzw. gegen Eins. Näheres zu diesem nur im Sinne der Probabilitätstheorie anwendbaren Ansatzes sind Höppner u. a. (1999, S. 49ff.) sowie Timm (2002, S 21f.) zu entnehmen.

Im Laufe der Zeit wurde noch eine Vielzahl weiterer Analyseverfahren zur Bestimmung unterschiedlichster Clusterformen wie z.B. das Linear-Clustering oder das Shell-Clustering entwickelt (vgl. z.B. Höppner u. a., 1999; Valente de Oliveira und Pedrycz, 2007), die jedoch aufgrund ihrer Festlegung auf bestimmte Clusterformen für die dynamische Analyse bei ökonomischen Fragestellungen von untergeordneter Bedeutung sind.

3.4. Possibilistische Clusteranalyse

In verschiedenen insbesondere ökomischen Szenarien zeigt sich die Wahrscheinlichkeit einer Clusterzugehörigkeit von untergeordneter Bedeutung; die Feststellung, inwieweit ein Objekt als *typisch* für die einzelnen Cluster anzusehen ist, erweist sich als wichtiger. Dies gilt im Besonderen, wenn sich die Cluster nicht gegenseitig ausschließen und ein Objekt tatsächlich zu mehreren Clustern gehören kann. Ein Beispiel hierfür bietet der Automobilmarkt: Dort gibt es unter anderem das Segment der umweltbewussten Autofahrer, die ökologische Aspekte favorisieren, und das der Familien, bei denen Sicherheit und Platzangebot große Bedeutung zukommt (vgl. Neumann, 2006, S. 61f.). Diese Segmente müssen sich jedoch nicht gegenseitig ausschließen: Ein Familienvater, der die sichere und bequeme Beförderung seiner Familie bevorzugt, kann gleichwohl an Umweltfaktoren interessiert sein. Bei Zuordnung im Sinne der probabilistischen Analyse würde er jedoch nur einem Cluster mit der höchsten Wahrscheinlichkeit zugeordnet werden; es ist sogar möglich, dass er aufgrund der Bedingung in Gleichung (3.3) zu beiden Clustern eine Zugehörigkeitswahrscheinlichkeit unter 0.5 bekäme und dadurch eine Zugehörigkeit zu einer der beiden Gruppen nicht verdeutlicht werden könnte. Auch bei der dynamischen Analyse darf diese Problematik nicht vernachlässigt werden: Bewegen sich zwei Kundensegmente über die Zeit hinweg aufeinander zu, so sollte dies anhand der wachsenden Zugehörigkeitsgrade der Kunden, die für diese Cluster typische Eigenschaftsausprägungen aufweisen, erkennbar sein. Unter Berücksichtigung der probabilistischen Bedingung in (3.3) würde diese Veränderung jedoch kaum deutlich werden, da die Zugehörigkeitsgrade immer entsprechend niedrig blieben. Das generelle Problem ist dabei in der Tatsache zu finden, dass die probabilistischen Zugehörigkeitsgrade die Datenstruktur nicht ausreichend repräsentieren können. Abbildung 3.5 soll diesen Umstand verdeutlichen: In Abbildung 3.5a sind fünf Objekte gegeben, die jeweils den gleichen Abstand zu den beiden Clusterzentren der durch Kreise gekennzeichneten Cluster besitzen, weshalb auch ihre Zugehörigkeitsgrade jeweils 0.5 betragen. Es ist nicht möglich, aus den Zugehörigkeitsgraden abzulesen, dass die einzelnen Objekte unterschiedlich weit von diesen Clustern entfernt, d.h., dass sie unterschiedlich *typisch* sind. In Abbildung 3.5b wird gezeigt, dass höhere Zugehörigkeitsgrade nicht zwangsläufig bedeuten, dass Objekte durch ein Cluster besser repräsentiert werden. Das Objekt zwischen den gezeigten Clustern bekommt wiederum Zugehörigkeitsgrade von jeweils 0.5, während die äußeren Objekte jeweils die gleiche Distanz zu einem der Cluster aufweisen, aber höhere Zugehörigkeitsgrade zu diesem näheren Cluster bekommen, da sie im Vergleich mit höherer Wahrscheinlichkeit diesem Cluster zugeordnet werden. Auch hier kann die Datenstruktur nicht aus den Zugehörigkeitsgraden abgeleitet werden.

(a) Fehlende Wiedergabe der Datenstruktur (b) Fehlende Wiedergabe der Clusterrepräsentativität

Abbildung 3.5.: Problematik der probabilistischen Analyse (vgl. Höppner u. a., 1999, S. 19)

Abhilfe bietet die possibilistische Clusteranalyse, die erstmals von Krishnapuram und Keller (1993) eingeführt wurde. Bei dieser Form der Clusteranalyse entfällt die Bedingung aus (3.3), so dass es möglich wird, ein Objekt mehreren Clustern zuzuordnen. Die Zugehörigkeitsgrade drücken also nicht die Wahrscheinlichkeit dafür aus, dass ein Objekt zu einem bestimmten Cluster gehört, sondern vielmehr die Möglichkeit, dass es diesem Cluster aufgrund seiner Charakteristika zuzuordnen ist, d.h. die *Typizität* der Objekte bzgl. der Cluster. Da ohne diese Nebenbedingung das Minimum der Zielfunktion in (3.1) jedoch in der trivialen Lösung läge, dass alle Zugehörigkeitsgrade Null oder unter Beachtung der Nebenbedingung in (3.2) verschwindend gering sind, wird bei der possibilistischen Analyse die Zielfunktion dahingehend erweitert, dass die einzelnen Zugehörigkeitsgrade möglichst hoch sein sollen, indem ein zweiter, hohe Zugehörigkeitsgrade belohnender Term eingeführt wird:

$$J_{poss}\left(X, U, C\right) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{C_i}^2\left(\vec{v}_i, \vec{x}_j\right) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(1 - u_{ij}\right)^m, \tag{3.9}$$

wobei

- X, U, C, m, $d_{C_i}^2\left(\vec{v}_i, \vec{x}_k\right)$: vgl. (3.1)
- η_i: Schätzung für die Clusterausdehnung; Objekte mit einer Distanz zum Clusterzentrum von $d_{C_i}^2\left(\vec{v}_i, \vec{x}_j\right) = \eta_i$ erhalten einen Zugehörigkeitsgrad von 0.5.

Der Parameter η_i legt die Ausdehnung eines Clusters fest. In (hyper-)sphärischen Clustern, wie sie vom Fuzzy-C-Means bestimmt werden, kann daraus direkt der Clusterdurchmesser $\sqrt{\eta_i}$ abgeleitet werden. Generell wird durch diesen Parameter für ein Cluster die Grenze festgelegt, ab der ein Objekt diesem Cluster mit einer Mindestzugehörigkeit von 0.5 zuzuordnen ist. Krishnapuram und Keller (1993) schlagen vor, die einzelnen η_i basierend auf einer zuvor durchgeführten probabilistischen Analyse zu schätzen, so dass die einem Cluster probabilistisch zugeordneten Objekte die Ausdehnung festlegen:

$$\eta_i = K\frac{\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2\left(\vec{v}_i, \vec{x}_j\right)}{\sum_{j=1}^{n} u_{ij}^m}, \tag{3.10}$$

wobei K ein Parameter zur Skalierung der Ausdehnung ist; i.d.R. wird $K = 1$ gewählt. Durch Minimierung der Zielfunktion in (3.9) ergibt sich damit Algorithmus 3.3 für den possibilistischen Fuzzy-C-Means zur Berechnung der Clusterprototypen $C_i = (\vec{v}_i, \eta_i)$. Wird anstelle des Fuzzy-C-Means der Ansatz nach Gustafson und Kessel (1979) verwendet, so liegen mit der Kovarianzmatrix Σ_i nähere Informationen bzgl. der Clusterausdehnung vor. In diesem Fall können die η_i direkt aus den Kovarianzmatrizen bestimmt werden:

$$\eta_i = \sqrt[p]{\det\left(\Sigma_i\right)} \tag{3.11}$$

Der possibilistische Gustafson-Kessel-Algorithmus für die Prototypen $C_i = (\vec{v}_i, \Sigma_i, \eta_i)$ wird in Algorithmus 3.4 dargestellt.

Krishnapuram und Keller (1993) empfehlen außerdem, für eine genauere Bestimmung des Parameters η_i diesen einmalig nach einigen Iterationen des possibilistischen Algorithmus erneut zu schätzen, sollte sich der Schätzwert auf Basis der probabilistischen Analyse als nicht ausreichend erweisen.

Algorithmus 3.3 Possibilistischer Fuzzy-C-Means (*possFCM*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix
- Initialisiere auf Basis der probabilistischen Analyse:
  - Clusterausdehnung η_i, $i = 1, \ldots, c$, nach (3.10),
  - Clusterprototypen $C_i^{(0)} = \left(\vec{v}_i^{probFCM}, \eta_i\right)$, $i = 1, \ldots, c$, und
  - Zugehörigkeitsgrade $U^{(0)} = U^{probFCM}$

repeat {für $\rho = 1, \ldots$}

Update der Distanzmatrix $D^{(\rho)} = \left(d^2_{C_i^{(\rho-1)}}(\vec{v}_i, \vec{x}_j)\right)$ bei konstanten $C_i^{(\rho-1)}$ mittels quadrierter euklidischer Distanz

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ mit

$$u_{ij} = \frac{1}{1 + \left(\frac{d^2_{C_i^{(\rho-1)}}(\vec{v}_i, \vec{x}_j)}{\eta_i}\right)^{\frac{1}{m-1}}} \tag{3.12}$$

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ nach (3.4)

Setze $C_i^{(\rho)} = (\vec{v}_i, \eta_i), i = 1, \ldots, c$

until $\left\|U^{(\rho-1)} - U^{(\rho)}\right\| < \epsilon$

Die Einbeziehung der Clusterausdehnung ermöglicht eine detailliertere Darstellung der Clusterstruktur, da nur noch die Distanz zum Clusterprototypen C_i für die Bestimmung des Zugehörigkeitsgrades zu ebendiesem Cluster relevant ist, nicht jedoch die Distanzen zu den übrigen Clustern. Dieses Vorgehen erweist sich insbesondere im Umgang mit Ausreißern und Stördaten als vorteilhaft, da es die Robustheit der Ergebnisse erhöht, beruhend auf der Tatsache, dass solche Objekte geringe Zugehörigkeiten zu allen Clustern vorweisen und somit insgesamt auf die Bestimmung der gesamten Clusterstruktur ein vergleichsweise geringes Gewicht erhalten. Allgemein lässt sich festhalten, dass zwar mit Hilfe einer probabilistische Analyse i.d.R. eine bessere Trennung der einzelnen Cluster, d.h. eine eindeutigere Partitionierung, erreicht wird, eine possibilistische Analyse jedoch eine bessere Abbildung der Datenstruktur bietet (vgl. Timm, 2002, S. 29f.).

Algorithmus 3.4 Possibilistischer Gustafson-Kessel-Algorithmus (*possGK*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix
- Initialisiere auf Basis der probabilistischen Analyse:
  - Clusterausdehnung η_i, $i = 1, \ldots, c$, nach (3.11),
  - Clusterprototypen $C_i^{(0)} = \left(\vec{v}_i^{probGK}, \eta_i, \Sigma_i^{probGK}\right), i = 1, \ldots, c$, und
  - Zugehörigkeitsgrade $U^{(0)} = U^{probFCM}$

repeat {für $\rho = 1, \ldots$}

Update der Distanzen $D^{(\rho)} = \left(d^2_{C_i^{(\rho-1)}}(\vec{v}_i, \vec{x}_j)\right)$ bei konstanten $C_i^{(\rho-1)}$ nach (3.7)

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ nach (3.12)

Update der Kovarianzmatrizen $\Sigma_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ und konstanten $\vec{v}_i$ gemäß $C_i^{(\rho-1)}$ nach (3.8)

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ nach (3.4)

Setze $C_i^{(\rho)} = (\vec{v}_i, \eta_i, \Sigma_i), i = 1, \ldots, c$

until $\left\|U^{(\rho-1)} - U^{(\rho)}\right\| < \epsilon$

3.5. Clustervalidität

Beim Fuzzy-Clustering handelt es sich um ein strukturentdeckendes Verfahren für ungelabelte Daten, das eine Evaluierung der ermittelten Clusterstruktur erfordert. Da i.d.R. in höherdimensionierten Räumen gearbeitet wird, reicht eine einfache Evaluierung in den meisten Fällen nicht aus, so dass sich das Hinzuziehen verschiedener Gütemaße als notwendig erweist. Anders als bei strukturprüfenden Verfahren wie der Klassifikation mit bekannten Klassenzugehörigkeiten existiert hier jedoch kein *optimales* Gütekriterium zur Ermittlung der besten Partitionierung (vgl. Abschnitt 1.2): Der Nutzer muss selbst entscheiden, welche Kriterien zur Beantwortung seiner jeweiligen Fragestellung geeignet sind; diese kann sich z.B. auf die gewählte Clusterzahl oder die Annahmen zur Clusterstruktur beziehen (vgl. Höppner u. a., 1999, S. 198).

Generell wird zwischen globalen und lokalen Gütemaßen unterschieden. Die letztgenannten überprüfen lediglich die Güte der einzelnen Cluster; sie werden insbesondere dann hinzugezogen, wenn in einer Clusterstruktur der Ort eventuell erforderlicher Nachbesserungen festgestellt werden soll (vgl. Höppner u. a., 1999, S. 20ff.; Timm, 2002, S. 36ff.). Den globalen Gütemaßen kommt dagegen eine größere Bedeutung zu, da sie die gesamte Clusterpartitionierung evaluieren. Die vier verbreitetesten globalen Gütemaße werden im Folgenden exemplarisch vorgestellt.

Partitionskoeffizient

Der Partitionskoeffizient nach Bezdek (1981, S. 100) beurteilt die Eindeutigkeit einer Clustereinteilung. Dabei werden diejenigen Clusterergebnisse bevorzugt, die möglichst harte Zugehörigkeitsgrade nahe Eins bzw. nahe Null erzielen.

$$PC\,(U) = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^2 \tag{3.13}$$

Je höher der Partitionskoeffizient ist, desto besser ist die Clustereinteilung zu beurteilen. Hierbei bleibt jedoch zu beachten, dass die Annahmen bzgl. der Interpretation der Zugehörigkeitsgrade beim Vergleich verschiedener Konfigurationen übereinstimmen müssen, um sinnvolle Ergebnisse zu erhalten. Am sinnvollsten wird der Partitionskoeffizient beim Vergleich verschiedener probabilistischer Zerlegungen eingesetzt.

Partitionsentropie

Die Partitionsentropie basiert auf Shannons Informationstheorie (vgl. Höppner u. a., 1999, S. 190). Wie der Name *Entropie* bereits impliziert, werden die Zugehörigkeitsgrade bei diesem Gütemaß als Informationsgehalt interpretiert; dementsprechend misst die Partitionsentropie den Unschärfegrad einer Zerlegung. Dabei werden wie beim Partitionskoeffizienten möglichst harte Zuordnungen bevorzugt; am ehesten empfiehlt sich deswegen eine Anwendung bei probabilistischen Clustereinteilungen. Die Partitionsentropie kann jedoch auch eingesetzt werden, um verschiedene Ergebnisse possibilistischer Verfahren zu vergleichen (vgl. Neumann, 2008, S. 62).

$$PE\,(U) = -\frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} \ln u_{ij} \tag{3.14}$$

Für eine harte Zerlegung nimmt die Partitionsentropie einen Wert von Null an. Im Allgemeinen gilt $PE \in [0, \ln c]$ (vgl. z.B. Cios u. a., 2007, S. 283); kleine Werte kennzeichnen gute Clusterergebnisse.

Separationsindex

Der Separationsindex nach Xie und Beni (1991) beschreibt die Fuzzy-Variante des aus dem harten Clustering bekannten Separationsindex nach Dunn (1973). Wie die bereits vorgestellten Validitätsmaße wurde er ursprünglich zur Evaluierung probabilistischer Clustereinteilungen entwickelt, kann jedoch ebenfalls auf possibilistische Ergebnisse angewandt werden. Der Separationsindex evaluiert die Trennung der einzelnen Cluster, indem die allgemeine Homogenität innerhalb der Cluster in Verhältnis gesetzt wird zur Heterogenität zwischen den Clusterprototypen.

$$S\,(U) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{C_i}^2\,(\vec{v}_i, \vec{x}_j)}{n \min\left\{d^2\,(C_i, C_{i'})\,|i, i' \in \{1, \ldots, c\}, i \neq i'\right\}} \tag{3.15}$$

Der Zähler in (3.15) repräsentiert die Homogenitätsbedingung unter Beachtung der Kompaktheit, der Nenner bezieht die Heterogenitätsbedingung ein. Folglich sind Ergebnisse mit einem kleinen Separationsindex zu bevorzugen. Die Distanz der Prototypen, $d^2\left(C_i, C_{i'}\right)$, wird allgemein als quadrierte euklidische Distanz bestimmt; sie kann jedoch problemlos auf ellipsoide Cluster und die Mahalanobisdistanz übertragen werden, indem der Mittelwert der Distanzen eines Clusterzentrums zum jeweils anderen Clusterprototypen ermittelt wird (vgl. Timm, 2002, S. 54):

$$d^2\left(C_i, C_{i'}\right) = \frac{1}{2}\left(d^2_{C_i}\left(\vec{v}_i, \vec{v}_{i'}\right) + d^2_{C_{i'}}\left(\vec{v}_{i'}, \vec{v}_i\right)\right).$$

Mittlere Partitionsdichte

Das Maß zur Evaluierung der mittleren Partitionsdichte (*A*verage *P*artition *D*ensity, *APD*) wurde von Gath und Geva (1989) eingeführt. Sie setzt die einem Cluster zugeordneten Daten, d.h. diejenigen Objekte, die eine maximale Distanz zum Clusterprototypen nicht überschreiten, in Verhältnis zum Volumen des Clusters, das aus der Determinante der Kovarianzmatrix bestimmt werden kann.

$$APD\left(U, C\right) = \frac{1}{c}\sum_{i=1}^{c}\frac{\mathrm{card}\left(Y_i\right)}{\sqrt{\det\left(\Sigma_i\right)}}, \tag{3.16}$$

wobei

- $Y_i = \left\{\vec{x}_j \middle| \left(\vec{x}_j - \vec{v}_i\right)^{\mathrm{T}} \Sigma_i^{-1} \left(\vec{x}_j - \vec{v}_i\right) < 1, j \in \{1, \ldots, n\}\right\}$,
- $\mathrm{card}(Y_i)$ – Fuzzy-Kardinalität von Y_i, d.h. $\mathrm{card}(Y_i) = \sum_{j=1}^{n} u_{ij}$.

Dieses Gütemaß kann auch als lokales Maß verwendet werden, indem die Dichte jedes Clusters separat bestimmt wird. Dadurch kann es im Kontext des Change Minings hinzugezogen werden, um Unterschiede in der Clusterstruktur aufzudecken (vgl. Kapitel 5). Je höher die mittlere Partitionsdichte ist, desto besser ist eine Clustereinteilung zu beurteilen.

A problem well-stated is a problem half-solved.

J. Dewey

4

Erweiterungen der possibilistischen Clusteranalyse

Wie in Abschnitt 3.4 bereits geschildert, erweist sich das possibilistische Clustering für die Clusteranalyse im Rahmen des Change Minings der probabilstischen Analyse oder dem harten Clustering als deutlich überlegen. Dies wird durch die Möglichkeit begründet, dass aufgrund der Interpretation der Zugehörigkeitsgrade als Typizitäten bzgl. der Cluster die Entwicklungen einzelner Cluster besser nachvollzogen werden. Begründen lässt sich dieser Umstand damit, dass es hohe Zugehörigkeitsgrade zu mehreren Clustern ebenso erlaubt wie das Erkennen von Ausreißern, die zu allen Clustern nur einen geringen Zugehörigkeitsgrad aufweisen. Ihre Anwendung beschränkt sich jedoch auf Fälle mit gut separierten Clustern, die zudem noch ungefähr dieselbe Größe und Form aufweisen müssen; bei Nichtbeachten dieser Problematik besteht die Gefahr der Clusteranziehung[12] (vgl. z.B. Borgelt, 2005, S. 54ff.; Timm, 2002, S. 45ff.). Auch wenn der in den genannten Quellen vorgestellte Beweis nur dann allgemeine Gültigkeit besitzt, wenn die η_i in der Zielfunktion (3.9) in jeder Iteration angepasst werden oder aber dieselben η_i für mehrere Cluster gelten (vgl. Abschnitt 4.2), darf dieses Problem nicht vernachlässigt werden: Sind zwei Cluster nicht ausreichend stark voneinander getrennt, erkennt die possibilistische Analyse nur ein Cluster, auch wenn sie intuitiv eindeutig voneinander zu differenzieren sind. Abbildung 4.1 stellt zwei nicht vollständig separierte Cluster dar: Intuitiv lassen sich diese Cluster aufgrund ihrer Ausrichtung und ihrer Lage zueinander gut unterscheiden, dennoch ist es auf Basis der possibilistischen Analyse nicht möglich, zwei Cluster mit einer solchen Nähe zueinander zu trennen.

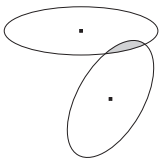


Abbildung 4.1.: Teilweise überlappende Cluster

Im Rahmen der Marktforschung kann dieser Umstand dazu führen, dass deutlich verschiedene Kundensegmente nicht ausreichend stark voneinander abgegrenzt werden und dadurch eine

[12] Bei der Bestimmung der Clusterprototypen handelt es sich um ein iteratives Vorgehen. Vollzieht man die in den einzelnen Iterationsschritten bestimmten Prototypen nach, wird bei schlecht separierten Clustern deutlich, dass diese sich sukzessive aufeinander zubewegen, d.h. sich gegenseitig anziehen.

eventuell notwendige getrennte Betrachtung verhindert wird. Ein weiteres, konkreteres Beispiel zu dieser Problematik wird in dem dieses Kapitel schließenden Abschnitt 4.4 zum experimentellen Vergleich der verschiedenen Ansätze gegeben.

Da sich die Bestimmung des Zeitpunktes, an dem eine Clusteranziehung auftritt, nicht nachvollziehen lässt, kann sich diese bei der Analyse dynamischer Veränderungen als Nachteil erweisen. Nähern sich z.B. zwei Cluster in der aktuellen Periode nach Perioden fehlender Clusteranziehung an, ohne dass sie als ein Cluster betrachtet werden dürfen, werden beide aufgrund der possibilistischen Analyse miteinander verschmolzen. Als Folge tritt ein nicht zu vernachlässigender Informationsverlust ein, da die Trennung in separate Cluster nicht länger erkennbar ist. Die Möglichkeit der Ausdehnung eines einzelnen Clusters mit nachfolgender Teilung besteht ebenfalls; in einer Analyse, die auf dem possibilistischen Clustering basiert, wird diese Veränderungen ebenfalls erst viel zu spät bemerkt.

Es gibt verschiedene Ansätze zur Anpassung der possibilistischen Analyse, mit denen dem Problem der Clusteranziehung begegnet werden kann; dieses Kapitel stellt verschiedene Techniken zur Lösung dieses Problems dar.

4.1. Kombination der probabilistischen und der possibilistischen Analyse

Zur Vermeidung der beschriebenen Clusteranziehung existieren verschiedene Möglichkeiten, den probabilistischen Ansatz zur Clusteranalyse mit possibilistischen Zugehörigkeitsgraden zu kombinieren; die einfachste besteht darin, mit Hilfe der probabilistischen Analyse, wie in Abschnitt 3.3 vorgestellt, die Clusterprototypen zu bestimmen. Anschließend werden die possibilistischen Zugehörigkeitsgrade gemäß der Updateregel in (3.12) bestimmt (vgl. z.B. Angstenberger, 2000, S. 83ff.). Dieses Vorgehen vermeidet die Nachteile der possibilistischen Clusteranalyse unter Anwendung des probabilistischen Vorgehens. Dieser einfache Ansatz birgt jedoch einen entscheidenen Nachteil: Bei der probabilistischen Analyse haben Ausreißer oder allgemein Stördaten einen – verglichen mit der possibilistischen Analyse – relativ hohen Einfluss auf die Bestimmung der Clusterprototypen und damit auf die resultierende Clusterstruktur (vgl. Abschnitt 3.4), da auch ihre Zugehörigkeitsgrade sich gemäß der Nebenbedingung in (3.3) zu Eins aufaddieren müssen. Die Ausprägung des konkreten Einflusses hängt dabei von den Zugehörigkeitsgraden sowie dem verwendeten Distanzmaß ab (vgl. Höppner u. a., 1999, S. 18f.).

Eine weitere Möglichkeit zur Kombination der verschiedenen Zugehörigkeitsarten besteht darin, eine kombinierte Zielfunktion zu verwenden, die sowohl probabilistische als auch possibilistische Zugehörigkeitsgrade enthält (vgl. z.B. Kruse u. a., 2007, S. 23ff.; Pal u. a., 2004). So gelingt es, die positiven Eigenschaften beider Arten an Zugehörigkeitsgraden einzubeziehen, d.h. die teilende Eigenschaft und damit die Vermeidung der Clusteranziehung bei der probabilistischen und die Reduzierung des Einflusses von Ausreißern bei der possibilistischen Analyse (vgl. Pal u. a., 1997):

$$J_{kombi}\left(X,U,C\right)=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\left(u_{ij}^{prob}\right)^{m_1}+\left(u_{ij}^{poss}\right)^{m_2}\right)d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right) \tag{4.1}$$

mit

- u_{ij}^{prob}: probabilistische Zugehörigkeitsgrade mit Fuzzifier m_1
- u_{ij}^{poss}: possibilistische Zugehörigkeitsgrade mit Fuzzifier m_2

unter Beachtung der Nebenbedingungen

$$\sum_{i=1}^{c} u_{ij}^{prob} = 1 \quad \forall j = 1, \ldots, n,$$
$$\sum_{j=1}^{n} u_{ij}^{poss} = 1 \quad \forall i = 1, \ldots, c.$$

Für die probabilistischen Zugehörigkeitsgrade u_{ij}^{prob} gilt gemäß den generellen Regeln der Wahrscheinlichkeitsrechnung weiterhin, dass die Zugehörigkeitsgrade für ein einzelnes Objekt in der Summe Eins ergeben müssen, bei den possibilistischen Zugehörigkeitsgraden u_{ij}^{poss} muss hingegen gelten, dass sich alle Zugehörigkeitsgrade eines einzelnen Clusters zu Eins aufaddieren lassen.

Die zugrunde liegende Zielfunktion (4.1) mit den genannten Nebenbedingungen für diesen Ansatz unterliegt jedoch dem Problem, dass eine allgemeine Gewichtung der unterschiedlichen Zugehörigkeitsgrade impliziert wird, da die probabilistischen Zugehörigkeitsgrade aufgrund der Skalierungsunterschiede dominieren (vgl. Kruse u. a., 2007, S. 24). Um dies zu vermeiden, führen Pal u. a. (2004) eine erweiterte Zielfunktion ein, die die Nebenbedingung betreffend der possibilistischen Zugehörigkeitsgrade vernachlässigt, fügen aber gleichzeitig analog zur originalen possibilistischen Zielfunktion (3.9) einen Term ein, der die triviale Lösung für die possibilistischen Zugehörigkeitsgrade u_{ij}^{poss} verhindern soll:

$$\begin{aligned} J'_{kombi}(X, U, C) &= \sum_{i=1}^{c}\sum_{j=1}^{n} \left(\gamma_{prob}\left(u_{ij}^{prob}\right)^{m_1} + \gamma_{poss}\left(u_{ij}^{poss}\right)^{m_2}\right) d_{C_i}^2(\vec{v}_i, \vec{x}_j) \\ &\quad + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} \left(1 - u_{ij}^{poss}\right)^{m_2} \end{aligned} \tag{4.2}$$

Die resultierenden Zugehörigkeitsgrade u_{ij}^{poss} können laut Kruse u. a. (2007, S. 25) jedoch nicht länger als Typizität von Objekten zu Clustern intepretiert werden, weshalb dieser Ansatz für die Betrachtung im Rahmen des Change Minings nicht geeignet ist.

Ein weiterer, auf Typizität basierender Ansatz wurde von Lesot und Kruse (2006) eingeführt. Im Zusammenhang mit der Analyse der dynamischen Entwicklung von Fuzzy-Clustern verfügt ihr Ansatz jedoch über geringen Nutzen, da die Interpretation der Zugehörigkeitsgrade, ein Objekt als *typisch* anzusehen, sich von der Interpretation unterscheidet, ein Objekt weise tatsächlich typische Eigenschaften verschiedener Cluster auf.

4.2. Modellierung der possibilistischen Analyse mit Hilfe der Clusterabstoßung

Um das geschilderte Problem der Clusteranziehung umgehen und dennoch die Interpretation der possibilistischen Zugehörigkeitsgrade als Typizität beibehalten zu können, erweitern Timm

u. a. (2001) den Originalansatz von Krishnapuram und Keller (1993) (vgl. Abschnitt 3.4). Die Erweiterung beruht auf der bereits einführend erwähnten Theorie, dass die Zielfunktion in (3.9) nur dann global minimiert wird, wenn alle Cluster sich vollständig überlagern (vgl. Borgelt, 2005, S. 54f.; Timm, 2002, S. 45ff.). Die Autoren argumentieren, dass die Cluster voneinander losgelöst zu betrachten seien: Die optimale Lage und Ausdehnung kann nach ihren Angaben für jedes Cluster separat bestimmt werden, indem die Zielfunktion aus (3.9) umgeformt wird zu

$$J_{poss}\left(X,U,C\right)=\sum_{i=1}^{c}\left(\sum_{j=1}^{n}u_{ij}^{m}d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_j\right)+\eta_i\sum_{j=1}^{n}\left(1-u_{ij}\right)^{m}\right).$$

Sie erörtern weiterhin, dass bei Gleichsetzung jedes Clusters mit dem der kleinsten Summe die Summe minimal wird. So zeigt Timm (2002, S. 46f.), dass für

$$\mathrm{sum}_i=\sum_{j=1}^{n}u_{ij}^{m}d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_j\right)+\eta_i\sum_{j=1}^{n}\left(1-u_{ij}^{m}\right)^{m},i=1,2,$$

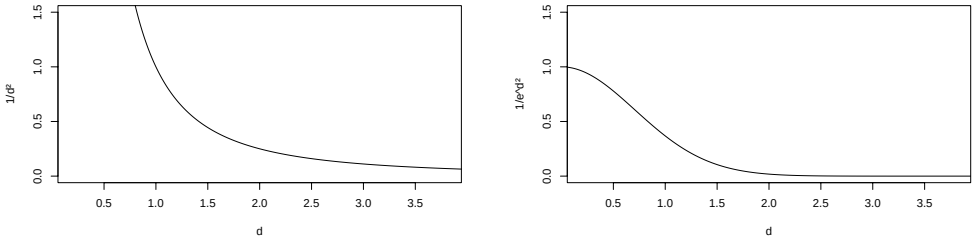
entweder $\mathrm{sum}_1<\mathrm{sum}_2$ oder $\mathrm{sum}_2<\mathrm{sum}_1$ gilt, wenn die Möglichkeit ausgeschlossen wird, dass eine sehr hohe Symmetrie im Datensatz vorherrscht. Sei o.B.d.A. $\mathrm{sum}_1<\mathrm{sum}_2$, dann gilt $\mathrm{sum}_1+\mathrm{sum}_1<\mathrm{sum}_1+\mathrm{sum}_2$, weshalb laut Timm die Zielfunktion nur dann global minimiert wird, wenn der Prototyp des zweiten Clusters dieselben Werte vorweist wie das erste Cluster. Diese Argumentation ist jedoch nur dann schlüssig, wenn alle η_i denselben Wert aufweisen oder aber die einzelnen η_i in jeder Iteration verändert werden, da durch sie die Gewichtung der Terme zueinander angepasst wird. Sind die η_i jedoch wie von Krishnapuram und Keller (1993) vorgeben (vgl. Abschnitt 3.4) abgesehen von einer einmalig möglichen Neuberechnung fixiert, verliert der Beweis an Gültigkeit. Auch wenn der Wert der Zielfunktion durch ein Gleichsetzen der Prototypen theoretisch verringert werden kann, ist diese Möglichkeit nicht gegeben: Sie würde ein nicht zulässiges Anpassen der Koeffizienten erfordern. Dessen ungeachtet bleibt die Anziehung der Cluster ein allgemeines Problem, wenn auch nicht in dem Ausmaß, das Timm (2002, S. 45ff.) und Borgelt (2005, S. 54f.) propagieren; erst bei unmittelbarer Nachbarschaft der Cluster zueinander tritt dieses Problem auf. Wie bereits einleitend erwähnt, kommt gerade im Rahmen des Change Minings dieser Anziehung entscheidende Bedeutung zu, da Trends sonst nicht oder nicht zum richtigen Zeitpunkt erkannt werden. Daher sind Algorithmen, die der Clusteranziehung entgegenwirken, für die Analyse dynamischer Veränderungen in der Clusterstruktur elementar.

Die Grundidee des erweiterten Ansatzes nach Timm u. a. (2001) liegt darin, in die Zielfunktion eine Heterogenitätsbedingung für Cluster einzubetten. Durch diese Bedingung sollen mit Hilfe einer sogenannten *Clusterabstoßung* Cluster daran gehindert werden, sich zu überlappen. Diese Bedingung wird in der originalen Zielfunktion (3.9) durch Einfühgen eines Bestrafungsterms modelliert, der die Distanzen zwischen den einzelnen Clusterprototypen einbezieht.

4.2.1. Der Bestrafungsterm

Die grundsätzliche Anforderung an den Bestrafungsterm besteht darin, dass er um so weniger schwerwiegend sein sollte, je weiter die Cluster voneinander entfernt sind (Timm, 2002, S. 48).

Für gut separierte Cluster sollte der Einfluss des Bestrafungsterms vernachlässigbar gering sein, dagegen im Fall von Clustern, die sich im Laufe des Verfahrens aufgrund der Anziehung sehr nahe kommen, die Zielfunktion dominieren. Auf Basis mathematischer Überlegungen führt Timm (2002, S. 49f.) zwei verschiedene Ansätze für den Bestrafungsterm ein: Die beiden in Abbildung 4.2 gegebenen Funktionen $f.\left(d^2\left(C_i, C_{i'}\right)\right)$ stellen diese Möglichkeiten dar, beide unter Einbeziehung der Distanzen zwischen den Clusterprototypen C_i und $C_{i'}$. Durch die Einbeziehung der Clusterdistanzen im Nenner der Funktionen wird die genannte Bedingung bzgl. des Einflusses des Bestrafungsterms auf die Zielfunktion erfüllt. Unter Anwendung von $f_1\left(d^2\left(C_i, C_{i'}\right)\right)$ ist die Abstoßung von benachbarten Clustern sehr stark (Abbildung 4.2a); im Gegensatz dazu ist $f_2\left(d^2\left(C_i, C_{i'}\right)\right)$ nach oben hin begrenzt (Abbildung 4.2b).



(a) $f_1\left(d^2\left(C_i, C_{i'}\right)\right) = \frac{1}{d^2(C_i, C_{i'})}$ (b) $f_2\left(d^2\left(C_i, C_{i'}\right)\right) = \frac{1}{e^{d^2(C_i, C_{i'})}}$

Abbildung 4.2.: Mögliche Funktionen für den Bestrafungsterm (vgl. Timm, 2002, S. 49)

Um den Bestrafungsterm in die Zielfunktion einzubinden, wird ferner ein clusterspezifischer Parameter γ_i benötigt, über den die jeweilige Funktion gewichtet wird; auf diese Weise erfolgt bei der Clusterabstoßung die Einbeziehung der Clustergröße. Dieser Parameter berechnet sich unter Verwendung der ermittelten Zugehörigkeitsgrade:

$$\gamma_i = \gamma \sum_{j=1}^{n} u_{ij}^m, \tag{4.3}$$

wobei $\gamma > 0$ ein allgemeiner, vorab festzulegender Gewichtungsparameter für den Bestrafungsterm ist. Ist $\gamma = 0$, liegt die originale Zielfunktion (3.9) nach Krishnapuram und Keller (1993) vor. Weiterhin führt Timm (2002, S. 49) einen Parameter ζ ein, über den die Prototypdistanzen normalisiert werden können; dieser Parameter hat einen Defaultwert von Eins. Je nach Art der Funktion, die für den Bestrafungsterm verwendet wird, ergeben sich die in (4.4) und (4.5) angegebenen Zielfunktionen.

$$J_{Het1}\left(X, U, C\right) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2\left(\vec{v}_i, \vec{x}_j\right) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(1 - u_{ij}\right)^m + \sum_{i=1}^{c}\gamma_i \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{1}{\zeta d^2\left(C_i, C_{i'}\right)} \tag{4.4}$$

$$J_{Het2}\left(X, U, C\right) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2\left(\vec{v}_i, \vec{x}_j\right) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(1 - u_{ij}\right)^m + \sum_{i=1}^{c}\gamma_i \sum_{\substack{i'=1 \\ i' \neq i}}^{c} e^{-\zeta d^2(C_i, C_{i'})} \tag{4.5}$$

4.2.2. Erweiterung des Algorithmus

Da in den Herleitungen der Updateregeln bei Timm (2002, S. 51ff.) Fehler auftreten, wurde eine Anpassung dieser vorgenommen, die ihrerseits zu einer Verbesserung der Ergebnisse im Vergleich zu den von Timm vorgestellten führte (vgl. Abschnitt 4.4). Da sich die Updateregeln, die im Folgenden Anwendung finden, von denen von Timm (2002, S. 52ff.) vorgestellten unterscheiden, werden die jeweiligen Algorithmen vollständig angegeben. Aufgrund der unterschiedlichen Distanzmaße sind sowohl die Alternativen für den Fuzzy-C-Means als auch für den Gustafson-Kessel-Algorithmus explizit aufgeführt. Die Herleitungen für die Updateregeln sind in Anhang A zu finden.

Im Falle des Fuzzy-C-Means sind die Clusterprototypen C_i wie in Abschnitt 3.4 eingeführt durch ihre Zentren $\vec{v}_i$ und ihre Ausdehnung η_i charakterisiert; aus diesem Grund ist die quadrierte euklidische Metrik als Distanzmaß $d^2\left(C_i, C_{i'}\right) = d^2\left(\vec{v}_i, \vec{v}_{i'}\right)$ für zwei Clusterprototypen C_i und $C_{i'}$ als symmetrisches Maß ausreichend. Die zugehörigen Algorithmen sind in Algorithmus 4.1 bzw. Algorithmus 4.2 aufgeführt.

Für den Gustafson-Kessel-Algorithmus stellt sich die Frage, welche Kovarianzmatrix für die Berechnung der Prototypdistanzen angewandt wird. Um ein symmetrisches Distanzmaß zu erhalten, schlägt Timm (2002, S. 54) vor, den Durchschnitt der einzelnen Distanzen aus Sicht des jeweiligen Clusters zu nutzen. Dabei werden anstelle der Kovarianzmatrizen die positiv definiten Normmatrizen $A_i = \det\left(\Sigma_i\right)^{\frac{1}{p}} \Sigma_i^{-1}$ verwendet:

$$\begin{aligned} d^2\left(C_i, C_{i'}\right) &= \tfrac{1}{2}\left(d^2_{C_i}\left(\vec{v}_i, \vec{v}_{i'}\right) + d^2_{C_{i'}}\left(\vec{v}_{i'}, \vec{v}_i\right)\right) \\ &= \tfrac{1}{2}\left(\left\|\vec{v}_{i'} - \vec{v}_i\right\|^2_{A_i} + \left\|\vec{v}_{i'} - \vec{v}_i\right\|^2_{A_{i'}}\right) \\ &= \tfrac{1}{2}\left(\left(\vec{v}_{i'} - \vec{v}_i\right)^{\mathrm{T}} A_i \left(\vec{v}_{i'} - \vec{v}_i\right) + \left(\vec{v}_{i'} - \vec{v}_i\right)^{\mathrm{T}} A_{i'} \left(\vec{v}_{i'} - \vec{v}_i\right)\right). \end{aligned} \tag{4.6}$$

Die Updateregeln zu den beiden Erweiterungen bzgl. der Clusterabstoßung sind in Algorithmus 4.3 und Algorithmus 4.4 angegeben. Dabei ist zu beachten, dass eine tatsächliche Berechnung der Kovarianzmatrix Σ_i für die einzelnen Cluster nicht möglich ist. Dieser Umstand liegt darin begründet, dass bei einer Minimierung der jeweiligen Zielfunktion hinsichtlich der Kovarianzmatrizen diese nicht mehr auf klassische Weise gemäß (3.8) berechnet werden können. Die Matrizen S_i, die durch eine Minimierung der Zielfunktionen bestimmt werden, können jedoch als Annäherung des Vielfachen der Kovarianzmatrizen betrachtet werden. Da eine Vervielfachung der Kovarianzmatrix die Normmatrix A_i nicht beeinflusst, wird für die Erweiterungen der Clusterprototyp dargestellt als $C_i = (\vec{v}_i, \eta_i, A_i)$.

Trotz vielversprechender Ergebnisse in einzelnen statischen Tests (vgl. Abschnitt 4.4) verfügt dieser Ansatz im Rahmen des Change Minings nur über einen begrenzten Wert, da er die Vereinigung der Cluster nach ausreichend starker Annäherung aneinander verhindert, ein bei der dynamischen Betrachtung einer Clusterstruktur durchaus vorstellbares Szenario. Am Beispiel von Marktsegmenten lässt sich das Problem verdeutlichen: Selbst wenn sich die charakterisierenden Eigenschaften verschiedener Kundensegmente über die Zeit hinweg anglichen, käme es weiterhin zu einer gegenseitigen Abstoßung ihrer Prototypen; als Folge wäre eine Annäherung nicht erkennbar und die Bearbeitung der Segmente erfolgte nach wie vor getrennt. Ein weiterer Nachteil liegt darin, dass die zwei Parameter γ und ζ stark von Größe und Nähe

der betrachteten Cluster abhängen, beides Eigenschaften, die sich im Laufe der Zeit verändern können; ihre dynamische Anpassung ist kaum möglich. Außerdem benötigt der Algorithmus von Timm u. a. (2001) einen hohen Rechenaufwand, da innerhalb einzelner Iterationen sowohl für die Clusterzentren $\vec{v}_i$ als auch – im Falle ellipsoider Cluster unter Anwendung des Gustafson-Kessel-Algorithmus – für die Normatrizen A_i wiederum iterative Berechnungen durchgeführt werden müssen.

Algorithmus 4.1 Fuzzy-C-Means mit Clusterabstoßung – Erweiterung 1 (*FCM-Het1*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix,
  - γ – allgemeiner Gewichtungsparameter,
  - ζ – Normalisierungsparameter.
- Initialisiere auf Basis des probabilistischen Fuzzy-C-Means:
  - η_i, $i = 1, \ldots, c$, nach (3.10),
  - $\gamma_i^{(0)}$, $i = 1, \ldots, c$, nach (4.3),
  - Clusterprotypen $C_i^{(0)} = \left(\vec{v}_i^{\,probFCM}, \eta_i\right)$, $i = 1, \ldots, c$,
  - Verteilung $U^{(0)} = U^{probFCM}$

repeat {für $\rho = 1, \ldots$}

Update der Distanzmatrix $D^{(\rho)} = \left(d^2_{C_i^{(\rho-1)}}\left(\vec{v}_i, \vec{x}_j\right)\right)$ bei konstanten $C_i^{(\rho-1)}$

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ nach (3.12)

Update der clusterspezifischen Gewichtungsparameter $\gamma_i^{(\rho)}, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ nach (4.3)

Setze $C_i^{(\rho)} = C_i^{(\rho-1)}, i = 1, \ldots, c$

repeat

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ und konstanten $\gamma_i^{(\rho)}$

$$\vec{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \vec{x}_j - \sum\limits_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}}{\zeta d^4\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)} \vec{v}_{i'}}{\sum_{j=1}^{n} u_{ij}^m - \sum\limits_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}}{\zeta d^4\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)}} \tag{4.7}$$

Setze $C_i^{(\rho)} = \left(\vec{v}_i, \eta_i\right), i = 1, \ldots, c$

until $\left\|\vec{v}_i^{(alt)} - \vec{v}_i^{(neu)}\right\| < \epsilon \; \forall i = 1, \ldots, c$

until $\left\|U^{(\rho-1)} - U^{(\rho)}\right\| < \epsilon$

Algorithmus 4.2 Fuzzy-C-Means mit Clusterabstoßung – Erweiterung 2 (*FCM-Het2*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix,
  - γ – allgemeiner Gewichtungsparameter,
  - ζ – Normalisierungsparameter.
- Initialisiere auf Basis des probabilistischen Fuzzy-C-Means:
  - η_i, $i = 1, \ldots, c$, nach (3.10),
  - $\gamma_i^{(0)}$, $i = 1, \ldots, c$, nach (4.3),
  - Clusterprotypen $C_i^{(0)} = \left(\vec{v}_i^{probFCM}, \eta_i\right), i = 1, \ldots, c$,
  - Verteilung $U^{(0)} = U^{probFCM}$

repeat {für $\rho = 1, \ldots$}

Update der Distanzmatrix $D^{(\rho)} = \left(d^2_{C_i^{(\rho-1)}}\left(\vec{v}_i, \vec{x}_j\right)\right)$ bei konstanten $C_i^{(\rho-1)}$

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ nach (3.12)

Update der clusterspezifischen Gewichtungsparameter $\gamma_i^{(\rho)}, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ nach (4.3)

Setze $C_i^{(\rho)} = C_i^{(\rho-1)} i = 1, \ldots, c$

repeat

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ und konstanten $\gamma_i^{(\rho)}$

$$\vec{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \vec{x}_j - \zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \left(\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}\right) e^{-\zeta d^2\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)} \vec{v}_{i'}}{\sum_{j=1}^{n} u_{ij}^m - \zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \left(\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}\right) e^{-\zeta d^2\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)}} \tag{4.8}$$

Setze $C_i^{(\rho)} = \left(\vec{v}_i, \eta_i\right), i = 1, \ldots, c$

until $\left\| \vec{v}_i^{(alt)} - \vec{v}_i^{(neu)} \right\| < \epsilon \; \forall i = 1, \ldots, c$

until $\left\| U^{(\rho-1)} - U^{(\rho)} \right\| < \epsilon$

Algorithmus 4.3 Gustafson-Kessel-Algorithmus mit Clusterabstoßung – Erweiterung 1 (*GK-Het1*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix,
  - γ – allgemeiner Gewichtungsparameter,
  - ζ – Normalisierungsparameter.
- Initialisiere auf Basis des probabilistischen Gustafson-Kessel-Algorithmus:
  - η_i, $i = 1, \ldots, c$, nach (3.11),
  - $\gamma_i^{(0)}$, $i = 1, \ldots, c$, nach (4.3),
  - Clusterprotypen $C_i^{(0)} = \left(\vec{v}_i^{probFCM}, \eta_i, A_i = \det\left(\Sigma_i^{probFCM}\right)^{\frac{1}{p}} \Sigma_i^{probFCM^{-1}}\right)$, $i = 1, \ldots, c$,
  - Verteilung $U^{(0)} = U^{probFCM}$

repeat {für $\rho = 1, \ldots$}

Update der Distanzmatrix $D^{(\rho)} = \left(d^2_{C_i^{(\rho-1)}}(\vec{v}_i, \vec{x}_j)\right)$ bei konstanten $C_i^{(\rho-1)}$

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ nach (3.12)

Update der clusterspezifischen Gewichtungsparameter $\gamma_i^{(\rho)}, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ nach (4.3)

Setze $C_i^{(\rho)} = C_i^{(\rho-1)}, i = 1, \ldots, c$

repeat

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ und konstanten $\gamma_i^{(\rho)}$ sowie A_i gemäß $C_i^{(\rho)}$

$$\begin{aligned}\vec{v}_i = &\left(\sum_{j=1}^{n} u_{ij}^m A_i^T - \sum_{\substack{i'=1\\ i'\neq i}}^{c} \frac{\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}}{2\zeta d^4\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)}\left(A_i^T + A_{i'}^T\right)\right)^{-1} \\ &\cdot \left(\sum_{j=1}^{n} u_{ij}^m \vec{x}_j^T A_i - \sum_{\substack{i'=1\\ i'\neq i}}^{c} \frac{\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}}{2\zeta d^4\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)}\left(\vec{v}_{i'}^T A_i + \vec{v}_{i'}^T A_{i'}\right)\right)^T\end{aligned} \tag{4.9}$$

Setze $C_i^{(\rho)} = (\vec{v}_i, \eta_i, A_i), i = 1, \ldots, c$

until $\left\|\vec{v}_i^{(alt)} - \vec{v}_i^{(neu)}\right\| < \epsilon \; \forall i = 1, \ldots, c$

repeat

Update der S_i und A_i, $i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ und konstanten $\gamma_i^{(\rho)}$ sowie $\vec{v}_i$ gemäß $C_i^{(\rho)}$

$$S_i = \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^T - \sum_{\substack{i'=1\\ i'\neq i}}^{c} \frac{\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}}{2\zeta d^4\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)}(\vec{v}_{i'} - \vec{v}_i)(\vec{v}_{i'} - \vec{v}_i)^T \tag{4.10}$$

$$A_i = \sqrt[p]{\det(S_i)} S_i^{-1}$$

Setze $C_i^{(\rho)} = (\vec{v}_i, \eta_i, A_i), i = 1, \ldots, c$

until $\left\|A_i^{(alt)} - A_i^{(neu)}\right\| < \epsilon \; \forall i = 1, \ldots, c$

until $\left\|U^{(\rho-1)} - U^{(\rho)}\right\| < \epsilon$

Algorithmus 4.4 Gustafson-Kessel-Algorithmus mit Clusterabstoßung – Erweiterung 2 (*GK-Het2*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix,
  - γ – allgemeiner Gewichtungsparameter,
  - ζ – Normalisierungsparameter.
- Initialisiere auf Basis des probabilistischen Gustafson-Kessel-Algorithmus:
  - η_i, $i = 1, \ldots, c$, nach (3.11),
  - $\gamma_i^{(0)}$, $i = 1, \ldots, c$, nach (4.3),
  - Clusterprotypen $C_i^{(0)} = \left(\vec{v}_i^{probFCM}, \eta_i, A_i = \det\left(\Sigma_i^{probFCM}\right)^{\frac{1}{p}} \Sigma_i^{probFCM^{-1}}\right)$, $i = 1, \ldots, c$,
  - Verteilung $U^{(0)} = U^{probFCM}$

repeat {für $\rho = 1, \ldots$}

Update der Distanzmatrix $D^{(\rho)} = \left(d^2_{C_i^{(\rho-1)}}\left(\vec{v}_i, ; \vec{x}_j\right)\right)$ bei konstanten $C_i^{(\rho-1)}$

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ nach (3.12)

Update der clusterspezifischen Gewichtungsparameter $\gamma_i^{(\rho)}, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ nach (4.3)

Setze $C_i^{(\rho)} = C_i^{(\rho-1)}, i = 1, \ldots, c$

repeat

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ und konstanten $\gamma_i^{(\rho)}$ sowie A_i gemäß $C_i^{(\rho)}$

$$\begin{aligned}\vec{v}_i = & \left(\sum_{j=1}^{n} u_{ij}^m A_i^T - \frac{1}{2}\zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \left(\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}\right) e^{-\zeta d^2\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)} \left(A_i^T + A_{i'}^T\right)\right)^{-1} \\ & \cdot \left(\sum_{j=1}^{n} u_{ij}^m \vec{x}_j^T A_i - \frac{1}{2}\zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \left(\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}\right) e^{-\zeta d^2\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)} \left(\vec{v}_{i'}^T A_i + \vec{v}_{i'}^T A_{i'}\right)\right)^T \end{aligned} \tag{4.11}$$

Setze $C_i^{(\rho)} = \left(\vec{v}_i, \eta_i, A_i\right), i = 1, \ldots, c$

until $\left\|\vec{v}_i^{(alt)} - \vec{v}_i^{(neu)}\right\| < \epsilon \forall i = 1, \ldots, c$

repeat

Update der S_i und A_i, $i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ und konstanten $\gamma_i^{(\rho)}$ sowie $\vec{v}_i$ gemäß $C_i^{(\rho)}$

$$S_i = \sum_{j=1}^{n} u_{ij}^m \left(\vec{x}_j - \vec{v}_i\right)\left(\vec{x}_j - \vec{v}_i\right)^T - \frac{1}{2}\zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \left(\gamma_i^{(\rho)} + \gamma_{i'}^{(\rho)}\right) e^{-\zeta d^2\left(C_i^{(\rho)}, C_{i'}^{(\rho)}\right)} \left(\vec{v}_{i'} - \vec{v}_i\right)\left(\vec{v}_{i'} - \vec{v}_i\right)^T$$

$$A_i = \sqrt[p]{\det\left(S_i\right)} S_i^{-1} \tag{4.12}$$

Setze $C_i^{(\rho)} = \left(\vec{v}_i, \eta_i, A_i\right), i = 1, \ldots, c$

until $\left\|A_i^{(alt)} - A_i^{(neu)}\right\| < \epsilon \ \forall i = 1, \ldots, c$

until $\left\|U^{(\rho-1)} - U^{(\rho)}\right\| < \epsilon$

4.3. Modellierung der possibilistischen Analyse unter Einbeziehung der Clusterhomogenität

Wie in Abschnitt 4.2 erläutert, eignet sich der dort beschriebene Ansatz nur bedingt für eine dynamische Analyse. Es wird daher ein Ansatz benötigt, der zwar die Clusteranziehung verringert, gleichzeitig jedoch die während der Untersuchung auftretenden Veränderungen darstellen kann. Der hierzu entwickelte Ansatz berücksichtigt weniger die Heterogentität zwischen den Clustern als vielmehr die Homogenität innerhalb der einzelnen Cluster, so dass tatsächlich nach Häufungspunkten gesucht wird (vgl. Minke u. a., 2009). Da es sich bei der Zielfunktion in (3.9) um eine zu minimierende Zielfunktion handelt, ist zu beachten, dass kleine Werte eine hohe Homogenität bedeuten.

4.3.1. Allgemeiner Ansatz

Um die Homogenität der von einem Cluster absorbierten Objekte zu bestimmen, wurde ein neuer Bestrafungsterm in die ursprüngliche possibilistische Zielfunktion (3.9) von Krishnapuram und Keller eingeführt. Dieser Bestrafungsterm enhält einen sogenannten α-*Schnitt* (vgl. Definition 4.2), so dass nur diejenigen Objekte darin einbezogen werden, die den einzelnen Clustern mit einem Mindestzugehörigkeitsgrad α zugeordnet werden; α wird dabei auch als *Grenzwert der Absorbierung* bezeichnet (vgl. Definition 4.1).

Definition 4.1. Sei ein Objekt j gegeben durch seinen Eigenschaftsvektor $\vec{x}_j$. Das Objekt wird von einem Cluster i mit dem Prototypen C_i *absorbiert*, falls für den Zugehörigkeitsgrad des Objekts j zum Cluster i gilt

$$u_{ij} \geq \alpha.$$

$\alpha \in [0, 1]$ heißt *Grenzwert der Absorbierung.*

Definition 4.2. (Kruse u. a., 1993, S. 16) Es sei $u \in F(O)$ und $\alpha \in [0, 1]$ der Grenzwert der Absorbierung. Dann heißt die Menge

$$[U]_\alpha := \{\vec{x}_j \in O | u_{\cdot j} \geq \alpha\}$$

der α-*Schnitt* von u.

Auf diese Weise beeinflussen die für ein Cluster nicht relevanten Objekte die Homogenitätsbestimmung dieses Clusters nicht. Da die Bewertung der Homogenität von der Größe der Cluster abhängt, wird auch bei diesem Ansatz ein clusterspezifischer Parameter γ_i verwendet (vgl. (4.13)), um eine Gewichtung des Bestrafungsterms abhängig von der Clustergröße zu erreichen. Dabei gilt, je größer ein Cluster ist, d.h. je größer η_i, desto höher dürfen auch die Objektdistanzen innerhalb des Clusters sein. Daher erfolgt die Gewichtung umgekehrt proportional zur Clustergröße:

$$\gamma_i = \frac{\gamma}{\eta_i}, \tag{4.13}$$

wobei γ ein allgemeiner Gewichtungsparameter mit einem Defaultwert von Eins ist. Neben der Clusterausdehnung ist jedoch auch die Anzahl der Objekte innerhalb des Clusters relevant,

daher erfolgt zusätzlich zur Gewichtung in Abhängigkeit von der Clustergröße eine Normierung auf Basis der Anzahl der durch ein Cluster absorbierten Objekte n_i^α. Zur Berechnung eines clusterspezifischen Normierungsparameters ζ_i sind verschiedene Möglichkeiten denkbar, z.B.

- $\zeta_i^{(1)} = \zeta n_i^\alpha$: Normierung in Abhängigkeit der Mächtigkeit von $[U_i]_\alpha$,
- $\zeta_i^{(2)} = \zeta \frac{n_i^\alpha \left(n_i^\alpha - 1\right)}{2}$: Durchschnittswert je Cluster,
- $\zeta_i^{(3)} = \zeta \frac{n_i^\alpha \left(n_i^\alpha - 1\right)}{2n}$: zusätzliche Einbeziehung der gesamten Objektzahl, da bei einer hohen Objektzahl ansonsten der Bestrafungsterm die Zielfunktion zu stark dominiert.

ζ ist ein allgemeiner Normierungsparameter mit einem Defaultwert von Eins. Die resultierende Zielfunktion ist in (4.14) gegeben.

$$J_{Hom}\left(X,U,C\right) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(1-u_{ij}\right)^m + \sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2\left(\vec{v}_i,\vec{x}_j,\vec{x}_k\right) \tag{4.14}$$

wobei

- $[U_i]_\alpha$: α-Schnitt bzgl. Cluster C_i,
- $d_{C_i}^2\left(\vec{v}_i,\vec{x}_j,\vec{x}_k\right)$ Distanzen bzgl. der Objekte $\vec{x}_j$ und $\vec{x}_k$ sowie dem Clusterzentrum $\vec{v}_i$ unter Berücksichtigung der Eigenschaften des Clusterprototypen C_i enthält.

Die Berechnung der Clusterhomogenität erfolgt nicht im klassischen Sinn unter Einbeziehung aller Distanzen der absorbierten Objekte untereinander; vielmehr ist bei der Berechnung auch die jeweilige Lage zum Clusterzentrum relevant.

Für den Distanzwert $d_{C_i}^2\left(\vec{v}_i,\vec{x}_j,\vec{x}_k\right)$ wurden verschiedene, im Folgenden näher erläuterte Ansätze entwickelt. Als Basis diente dabei die einfache These, dass der Distanzwert als durchschnittliche Distanz zweier Objekte zum Clustermittelpunkt bestimmt wird:

$$d_{C_i\,HomB}^2\left(\vec{v}_i,\vec{x}_j,\vec{x}_k\right) = \frac{1}{2}\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right) + d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right) \tag{4.15}$$

Dadurch erigbt sich als neue Zielfunktion

$$\begin{aligned} J_{HomB}\left(X,U,C\right) = \; & \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right) + \sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}\left(1-u_{ij}\right)^m \\ & + \frac{1}{2}\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right) + d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right). \end{aligned} \tag{4.16}$$

Die resultierenden Updateregeln können aus Algorithmus 4.5 entnommen werden, die Herleitungen zu diesem und den übrigen Ansätzen zur Modellierung der Clusterhomogenität sind in Anhang B aufgeführt. Es ist zu beachten, dass auch bei den hier eingeführten Erweiterungen anstelle der Kovarianzmatrizen Σ_i die Annäherungen S_i bestimmt werden (vgl. Abschnitt 4.2.2); die Clusterprototypen ergeben sich somit zu $C_i = (\vec{v}_i, \eta_i, A_i)$.

Algorithmus 4.5 Basisansatz zur Modellierung der Clusterhomogenität (*FCM-HomB* bzw. *GK-HomB*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix,
  - γ – allgemeiner Gewichtungsparameter,
  - ζ – Normalisierungsparameter.
- Initialisiere auf Basis der probabilistischen Analyse:
  - η_i, $i = 1, \ldots, c$, nach (3.10) bzw. nach (3.11),
  - γ_i, $i = 1, \ldots, c$, nach (4.13),
  - $\zeta_i^{(0)}$, $i = 1, \ldots, c$,
  - Clusterprotypen $C_i^{(0)} = \left(\vec{v}_i^{\,probFCM}, \eta_i\right)$
    bzw. $C_i^{(0)} = \left(\vec{v}_i^{\,probFCM}, \eta_i, A_i = \det\left(\Sigma_i^{probFCM}\right)^{\frac{1}{p}} \Sigma_i^{probFCM^{-1}}\right)$, $i = 1, \ldots, c$,
  - Verteilung $U^{(0)} = U^{probFCM}$

repeat {für $\rho = 1, \ldots$}

Update der Distanzmatrix $D^{(\rho)} = \left(d^2_{C_i^{(\rho-1)}}\left(\vec{v}_i, \vec{x}_j\right)\right)$ bei konstanten $C_i^{(\rho-1)}$

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ nach (3.12)

Update der clusterspezifischen Normierungsparameter $\zeta_i^{(\rho)}, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ und konstanten $\zeta_i^{(\rho)}$

$$\vec{v}_i = \frac{\displaystyle\sum_{j=1}^{n} u_{ij}^m \vec{x}_j + \frac{\gamma_i}{2\zeta_i^{(\rho)}} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \left(\vec{x}_j + \vec{x}_k\right)}{\displaystyle\sum_{j=1}^{n} u_{ij}^m + \frac{\gamma_i}{2\zeta_i^{(\rho)}} n_i^\alpha \left(n_i^\alpha - 1\right)}, \tag{4.17}$$

Im Falle des Gustafson-Kessel: Update der S_i und $A_i, i = 1, \ldots, c$, bei konstanten $\vec{v}_i$ gemäß $C_i^{(\rho-1)}$

$$\begin{aligned} S_i \;=\; & \sum_{j=1}^{n} u_{ij}^m \left(\vec{x}_j - \vec{v}_i\right)\left(\vec{x}_j - \vec{v}_i\right)^T \\ & + \frac{\gamma_i}{2\zeta_i^{(\rho)}} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \left(\left(\vec{x}_j - \vec{v}_i\right)\left(\vec{x}_j - \vec{v}_i\right)^T + \left(\vec{x}_k - \vec{v}_i\right)\left(\vec{x}_k - \vec{v}_i\right)^T\right) \end{aligned} \tag{4.18}$$

$$A_i = \sqrt[p]{\det\left(S_i\right)} S_i^{-1}$$

Setze $C_i^{(\rho)} = \left(\vec{v}_i, \eta_i\right)$ bzw. $C_i^{(\rho)} = \left(\vec{v}_i, \eta_i, A_i\right), i = 1, \ldots, c$

until $\left\|U^{(\rho-1)} - U^{(\rho)}\right\| < \epsilon$

Der Vorteil dieses Ansatzes gegenüber den in den folgenden Abschnitten vorgestellten besteht darin, dass aufgrund seiner Einfachheit die Berechnung wenig zeitintensiv ist. Nachteilig wirkt sich jedoch aus, dass die tatsächliche Distanz zwischen den einzelnen Objekten nicht berücksichtigt wird und die erreichten Ergebnisse oftmals über eine geringere Genauigkeit verfügen (vgl. Neumann, 2008, S. 30ff.).

4.3.2. Verwendung des Verhältnisses von Distanzen

Eine andere Option zur Bestimmung der $d^2_{C_i}(\vec{v}_i, \vec{x}_j, \vec{x}_k)$ besteht in der Verwendung des Verhältnisses der Distanzen untereinander. Diese Möglichkeit erscheint intuitiv sinnvoller als die Verwendung der durchschnittlichen Distanz der Objekte zum Clusterzentrum, da sich dabei eines der beiden folgenden zwei Ziele verfolgen lässt:

1. Generelle Minimierung der Distanzen innerhalb eines Clusters,
2. Annäherung des Verhältnisses der Distanzen an den Wert 1.

Im ersten Fall kann die durchschnittliche Distanz innerhalb eines Clusters bzw. die der aktuell betrachteten Objekte unter Beachtung der Distanz zwischen den einzelnen Objekten minimiert werden: Objekte, die weit voneinander entfernt liegen, besitzen nicht dieselbe Relevanz für die Positionierung des Clusterzentrums wie solche, die nah aneinander liegen; dieses Vorgehen führt zu der in (4.19) gegebenen Distanzfunktion.

$$d^2_{C_i HomV_1}(\vec{v}_i, \vec{x}_j, \vec{x}_k) = \frac{\frac{1}{2}\left(d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)\right)}{d^2_{C_i}(\vec{x}_j, \vec{x}_k)} \tag{4.19}$$

Um den Fall identischer Objekte ($d^2_{C_i}(\vec{x}_j, \vec{x}_k) = 0$) und der damit verbundenen Division durch Null entgegenzuwirken, wird bei einem solchen Fall die Distanz auf einen vernachlässigbar kleinen Wert μ gesetzt.[13]

Der Ansatz führt zu einer erhöhten Wahrscheinlichkeit, Häufungspunkte innerhalb des Clusters anstelle des tatsächlichen Clusterzentrums zu bestimmen. Eine Umkehrung der Relation führt zu der Minimierung der Objektdistanz im Vergleich zu ihrer durchschnittlichen Distanz (vgl. (4.20)) bzw. der durchschnittlichen Distanz aller Objekte zum Clusterzentrum (vgl. (4.21)). Dieses Vorgehen erscheint sinnvoller, da eine Erhöhung der durchschnittlichen Distanz innerhalb des Clusters zwangsläufig zu einer größeren Entfernung der vom Cluster absorbierten Objekte führen kann.

$$d^2_{C_i HomV_2}(\vec{v}_i, \vec{x}_j, \vec{x}_k) = \frac{d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{\frac{1}{2}\left(d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)\right)} \tag{4.20}$$

$$d^2_{C_i HomV_3}(\vec{v}_i, \vec{x}_j, \vec{x}_k) = \frac{d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{\frac{1}{n_i^\alpha}\sum_{\vec{x}_l \in [U_i]_\alpha} d^2_{C_i}(\vec{v}_i, \vec{x}_l)} \tag{4.21}$$

[13] Dieser Fall ist nicht erwünscht, da laut der Identitätsbedingung einer Metrik gilt $d^2_{C_i}(\vec{x}_j, \vec{x}_k) = 0 \Rightarrow j = k$. In großen Datensätzen kann dies jedoch aufgrund der Reduktion auf wenige Eigenschaften nicht ausgeschlossen werden.

Durch Einfügen von (4.20) bzw. (4.21) in die allgemeine Zielfunktion (4.14) erhält man die in (4.22) bzw. (4.23) gegebenen Zielfunktionen.

$$\begin{aligned} J_{HomV_2}(X,U,C) = & \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2(\vec{v}_i,\vec{x}_j) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}(1-u_{ij})^m \\ & +2\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} \frac{d_{C_i}^2(\vec{x}_j,\vec{x}_k)}{d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)} \end{aligned} \tag{4.22}$$

$$\begin{aligned} J_{HomV_3}(X,U,C) = & \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2(\vec{v}_i,\vec{x}_j) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}(1-u_{ij})^m \\ & +\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}n_i^\alpha \frac{\displaystyle\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2(\vec{x}_j,\vec{x}_k)}{\displaystyle\sum_{\vec{x}_j\in[U_i]_\alpha} d_{C_i}^2(\vec{v}_i,\vec{x}_j)} \end{aligned} \tag{4.23}$$

Da sich die beiden Algorithmen stark ähneln, wird im Folgenden nur der zu (4.22) gehörige Algorithmus 4.6 angegeben.

Die zweite Bedingung, dass das Verhältnis zwischen den Distanzen möglichst Eins sein soll, basiert auf der Idee, dass das Clusterzentrum tatsächlich mittig zwischen den von dem Cluster absorbierten Objekten liegen soll. Basierend auf der in (4.20) gegeben Distanzfunktion werden die Distanzen innerhalb des Clusters unter Anwendung der Distanzfunktion

$$d_{C_i\,HomV_4}^2(\vec{v}_i,\vec{x}_j,\vec{x}_k) = \left(1-\frac{2d_{C_i}^2(\vec{x}_j,\vec{x}_k)}{d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)}\right)^2 \tag{4.26}$$

berechnet. Diese Distanzfunktion führt zu der Zielfunktion

$$\begin{aligned} J_{HomV_4}(X,U,C) = & \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2(\vec{v}_i,\vec{x}_j) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}(1-u_{ij})^m \\ & +\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} \left(1-\frac{2d_{C_i}^2(\vec{x}_j,\vec{x}_k)}{d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)}\right)^2 . \end{aligned} \tag{4.27}$$

Die Nutzung des Verhältnisses zwischen Objektdistanz und durchschnittlicher Distanz zum Clusterzentrum erscheint durchaus plausibel, dennoch verfügt sie über einen gravierenden Nachteil: Obwohl unter Anwendung der quadrierten euklidischen Distanz und des Fuzzy-C-Means durchaus gute Ergebnisse erzielt werden konnten (vgl. Neumann, 2008, S. 27), kann dieser Ansatz nur sehr bedingt unter Verwendung der Mahalanobisdistanz angewandt werden. Die Ursache dafür liegt darin, dass für zu hohe Werte von γ, d.h. für eine hohe Gewichtung des Bestrafungsterms, der Bestrafungsterm die Zielfunktion derartig dominiert, dass die Determinante von S_i negativ wird und so eine Berechnung der positiv definiten Normmatrix nicht länger möglich ist; dieses Problem kann auftreten, da es sich bei den S_i lediglich um Annäherungen des Vielfachen der Kovarianzmatrix handelt, bei deren Schätzungen der Bestrafungsterm einbezogen wird. Für zu kleine Werte für γ ist der Bestrafungsterm jedoch zu schwach, um

Algorithmus 4.6 Verhältnis der Distanzen zur Modellierung der Clusterhomogenität (*FCM-HomV2* bzw. *GK-HomV2*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix,
  - γ – allgemeiner Gewichtungsparameter,
  - ζ – Normalisierungsparameter.
- Initialisiere auf Basis der probabilistischen Analyse:
  - η_i, $i = 1, \ldots, c$, nach (3.10) bzw. nach (3.11),
  - γ_i, $i = 1, \ldots, c$, nach (4.13),
  - $\zeta_i^{(0)}$, $i = 1, \ldots, c$,
  - Clusterprotypen $C_i^{(0)} = \left(\vec{v}_i^{\,probFCM}, \eta_i\right)$
    bzw. $C_i^{(0)} = \left(\vec{v}_i^{\,probFCM}, \eta_i, A_i = \det\left(\Sigma_i^{probFCM}\right)^{\frac{1}{p}} \Sigma_i^{probFCM^{-1}}\right)$, $i = 1, \ldots, c$,
  - Verteilung $U^{(0)} = U^{probFCM}$

repeat {für $\rho = 1, \ldots$}

Update der Distanzmatrix $D^{(\rho)} = \left(d^2_{C_i^{(\rho-1)}}(\vec{v}_i, \vec{x}_j)\right)$ bei konstanten $C_i^{(\rho-1)}$

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ nach (3.12)

Update der clusterspezifischen Normierungsparameter $\zeta_i^{(\rho)}, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$

Setze $C_i^{(\rho)} = C_i^{(\rho-1)}, i = 1, \ldots, c$

repeat

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstanten $U^{(\rho)}$, $D^{(\rho)}$ und konstanten $\zeta_i^{(\rho)}$

$$\vec{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \vec{x}_j - 2\frac{\gamma_i}{\zeta_i^{(\rho)}} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \frac{d^2_{C_i^{(\rho)}}(\vec{x}_j, \vec{x}_k)}{\left(d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{x}_j) + d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{x}_k)\right)^2} (\vec{x}_j + \vec{x}_k)}{\sum_{j=1}^{n} u_{ij}^m - 4\frac{\gamma_i}{\zeta_i^{(\rho)}} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \frac{d^2_{C_i^{(\rho)}}(\vec{x}_j, \vec{x}_k)}{\left(d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{x}_j) + d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{x}_k)\right)^2}} \tag{4.24}$$

Setze $C_i^{(\rho)} = (\vec{v}_i, \eta_i, A_i), i = 1, \ldots, c$; Update $D^{(\rho)}$ (s.o.)

until $\left\|\vec{v}_i^{(alt)} - \vec{v}_i^{(neu)}\right\| < \epsilon \; \forall i = 1, \ldots, c$

repeat

Im Falle des Gustafson-Kessel-Algrotihmus: Update der S_i und $A_i, i = 1, \ldots, c$, bei konstanten $U^{(\rho)}$, $D^{(\rho)}$ und konstanten $\zeta_i^{(\rho)}$ sowie $\vec{v}_i$ gemäß $C_i^{(\rho)}$

$$\begin{aligned} S_i = & \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^T + 2\frac{\gamma_i}{\zeta_i^{(\rho)}} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \left(\frac{1}{\left(d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{x}_j) + d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{x}_k)\right)^2}\right. \\ & \cdot \left(\left(d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{x}_j) + d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{x}_k)\right)(\vec{x}_j - \vec{x}_k)(\vec{x}_j - \vec{x}_k)^T\right. \\ & \left.\left. - d^2_{C_i^{(\rho)}}(\vec{x}_j, \vec{x}_k)\left((\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^T + (\vec{x}_k - \vec{v}_i)(\vec{x}_k - \vec{v}_i)^T\right)\right)\right) \\ A_i = & \sqrt[p]{\det(S_i)} S_i^{-1} \end{aligned} \tag{4.25}$$

Setze $C_i^{(\rho)} = (\vec{v}_i, \eta_i, A_i), i = 1, \ldots, c$; Update $D^{(\rho)}$ (s.o.)

until $\left\|A_i^{(alt)} - A_i^{(neu)}\right\| < \epsilon \; \forall i = 1, \ldots, c$

until $\left\|U^{(\rho-1)} - U^{(\rho)}\right\| < \epsilon$

dem allgemeinen Problem der Clusteranziehung entgegenzuwirken. Das mögliche Intervall für γ hängt dabei von verschiedenen Parametern ab, unter anderem von der Lage der Cluster zueinander, von ihrer Form und von der Distanz zwischen den absorbierten Objekten eines Clusters, und kann sich zwischen zwei Iterationen ändern. Ähnlich verhält es sich bzgl. des Einflusses des Normierungsparameters ζ_i: Je stärker der Bestrafungsterm im Vergleich zu den übrigen Termen der Zielfunktion ist, desto eher tritt das Problem der negativen Determinante von S_i auf. Auch hier ist eine geeignete Wahl von der Clusterlage und der Struktur der einzelnen Cluster abhängig. Daher ist es kaum möglich, einen geeigneten Gewichtungsparameter γ sowie ein geeignetes ζ_i für eine einzelne Periode, d.h. für eine statische Betrachtung, vorzugeben, noch weniger für mehrere aufeinanderfolgende Betrachtungen. Eine mögliche Lösung dieses Problems bietet die Berechnung der Kovarianzmatrix gemäß (3.8) ohne Beachtung der Zielfunktion, d.h., indem lediglich die Distanzen zum Clusterzentrum einbezogen werden (vgl. (3.8)).

4.3.3. Verwendung einer Dreiecksbeziehung der Distanzen

Ein weiterer, vielversprechende Ergebnisse liefernder Ansatz (vgl. Abschnitt 4.4) basiert nicht auf dem Verhältnis der Distanzen zwischen Objekten und Clusterzentrum, sondern auf dem Produkt aus der Objektdistanz zwischen Objekt j und Objekt k und ihrer durchschnittlichen Distanz zum Clusterzentrum (vgl. Minke u. a., 2009):

$$d^2_{C_i\,HomD}\left(\vec{v}_i, \vec{x}_j, \vec{x}_k\right) = \frac{1}{2} d^2_{C_i}\left(\vec{x}_j, \vec{x}_k\right)\left(d^2_{C_i}\left(\vec{v}_i, \vec{x}_j\right) + d^2_{C_i}\left(\vec{v}_i, \vec{x}_k\right)\right) \tag{4.28}$$

Der Kern dieses Ansatzes liegt in der Bestimmung des tatsächlichen Clustermittelpunktes anstelle von Häufungspunkten eines Clusters. Da alle in den Bestrafungsterm einbezogenen Objekte mit einem Mindestzugehörigkeitsgrad von α von dem jeweils betrachteten Cluster absorbiert werden (vgl. Abschnitt 4.3.1), besitzt jedes Relevanz für die Bestimmung des Prototypen. Um das Clusterzentrum in die Mitte des Clusters zu ziehen, erhalten die am weitesten voneinander entfernt, d.h. die an den Clustergrenzen liegenden Objekte das höchste Gewicht. Außerdem wird durch die Verwendung des Produktes anstelle des Quotienten das allgemeine Ziel erreicht, die Distanzen zwischen Objekten und Clusterzentrum zu minimieren. Dies bedeutet einen entscheidenden Vorteil, da gerade in Marktforschungsproblemen nicht perfekt normalverteilte Cluster, sondern häufig verschiedene Gebiete mit leicht erhöhter Dichte vorliegen, ohne dass die Cluster zu unterteilen wären.

Durch Einsetzen von (4.28) in (4.14) erhält man

$$\begin{aligned} J_{HomD}\left(X, U, C\right) = & \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d^2_{C_i}\left(\vec{v}_i, \vec{x}_j\right) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(1-u_{ij}\right)^m \\ & + \frac{1}{2}\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d^2_{C_i}\left(\vec{x}_j, \vec{x}_k\right)\left(d^2_{C_i}\left(\vec{v}_i, \vec{x}_j\right) + d^2_{C_i}\left(\vec{v}_i, \vec{x}_k\right)\right) \end{aligned} \tag{4.29}$$

Die zugehörigen Updateregeln sind in Algorithmus 4.7 angegeben.

Algorithmus 4.7 Dreiecksbeziehung der Distanzen zur Modellierung der Clusterhomogenität (*FCM-HomD* bzw. *GK-HomD*)

Initialisierung

- Gegeben:
  - c – Anzahl Cluster,
  - m – Fuzzifier,
  - $X = (x_{jl})$ – Datenmatrix,
  - γ – allgemeiner Gewichtungsparameter,
  - ζ – Normalisierungsparameter.
- Initialisiere auf Basis der probabilistischen Analyse:
  - η_i, $i = 1, \ldots, c$, nach (3.10) bzw. nach (3.11),
  - γ_i, $i = 1, \ldots, c$, nach (4.13),
  - $\zeta_i^{(0)}$, $i = 1, \ldots, c$,
  - Clusterprotypen $C_i^{(0)} = \left(\vec{v}_i^{\,probFCM}, \eta_i\right)$
    bzw. $C_i^{(0)} = \left(\vec{v}_i^{\,probFCM}, \eta_i, A_i = \det\left(\Sigma_i^{probFCM}\right)^{\frac{1}{p}} \Sigma_i^{probFCM^{-1}}\right)$, $i = 1, \ldots, c$,
  - Verteilung $U^{(0)} = U^{probFCM}$

repeat {für $\rho = 1, \ldots$}

Update der Distanzmatrix $D^{(\rho)} = \left(d^2_{C_i^{(\rho-1)}}(\vec{v}_i, \vec{x}_j)\right)$ bei konstanten $C_i^{(\rho-1)}$

Update der Zugehörigkeitsgrade $U^{(\rho)} = (u_{ij})$ bei konstantem $D^{(\rho)}$ nach (3.12)

Update der clusterspezifischen Normierungsparameter $\zeta_i^{(\rho)}, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$

Update der Clusterzentren $\vec{v}_i, i = 1, \ldots, c$, bei konstantem $U^{(\rho)}$ und konstanten $\zeta_i^{(\rho)}$

$$\vec{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \vec{x}_j + \frac{\gamma_i}{2\zeta_i^{(\rho)}} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} d^2_{C_i^{(\rho-1)}}(\vec{x}_j, \vec{x}_k)(\vec{x}_j + \vec{x}_k)}{\sum_{j=1}^{n} u_{ij}^m + \frac{\gamma_i}{\zeta_i^{(\rho)}} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} d^2_{C_i^{(\rho-1)}}(\vec{x}_j, \vec{x}_k)} \tag{4.30}$$

Setze $C_i^{(\rho)} = (\vec{v}_i, \eta_i, A_i), i = 1, \ldots, c$; Update $D^{(\rho)}$ (s.o.)

repeat

Im Falle des Gustafson-Kessel-Algrotihmus: Update der S_i und $A_i, i = 1, \ldots, c$, bei konstanten $U^{(\rho)}$, $D^{(\rho)}$ und konstanten $\zeta_i^{(\rho)}$ sowie $\vec{v}_i$ gemäß $C_i^{(\rho)}$

$$\begin{aligned} S_i = & \sum_{j=1}^{n} u_{ij}^m (\vec{x}_k - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^T \\ & + \frac{\gamma_i}{2\zeta_i^{(\rho)}} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \Bigg(\left(d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{x}_j) + d^2_{C_i^{(\rho)}}(\vec{v}_i, \vec{v}_k) \right) (\vec{x}_j - \vec{x}_k)(\vec{x}_j - \vec{x}_k)^T \\ & + d^2_{C_i^{(\rho)}}(\vec{x}_j, \vec{x}_k) \left((\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^T + (\vec{x}_k - \vec{v}_i)(\vec{x}_k - \vec{v}_i)^T \right) \Bigg) \end{aligned} \tag{4.31}$$

$$A_i = \sqrt[p]{\det(S_i)} S_i^{-1}$$

Setze $C_i^{(\rho)} = (\vec{v}_i, \eta_i, A_i), i = 1, \ldots, c$; Update $D^{(\rho)}$ (s.o.)

until $\left\| A_i^{(alt)} - A_i^{(neu)} \right\| < \epsilon \; \forall i = 1, \ldots, c$

until $\left\| U^{(\rho-1)} - U^{(\rho)} \right\| < \epsilon$

4.4. Experimenteller Vergleich

Zum Vergleich des Ansatzes von Timm (Abschnitt 4.2) und der auf der Homogenität der Cluster basierenden Ansätze (Abschnitt 4.3) mit dem originalen possibilistischen, der in Abschnitt 3.4 näher beschrieben wurde, sowie der probabilistischen Analyse (Abschnitt 3.3) wurden die Algorithmen auf verschiedenen Datensätzen getestet. Dieser Abschnitt stellt exemplarisch die Ergebnisse des bekannten Weindatensatzes[14] (Asuncion und Newman, 2007) dar; dabei werden aufgrund der besseren optischen Nachvollziehbarkeit nur die Eigenschaften 7 (Flavonoide) und 10 (Farbintensität) für die Analyse verwendet. Zunächst erfolgt eine grafische Darstellung der Ergebnisse, anschließend wird ein Vergleich der Algorithmen anhand verschiedener Validitätsmaße unter Berücksichtigung alternativer α-Werte vorgenommen.

4.4.1. Grafische Darstellung der Ergebnisse

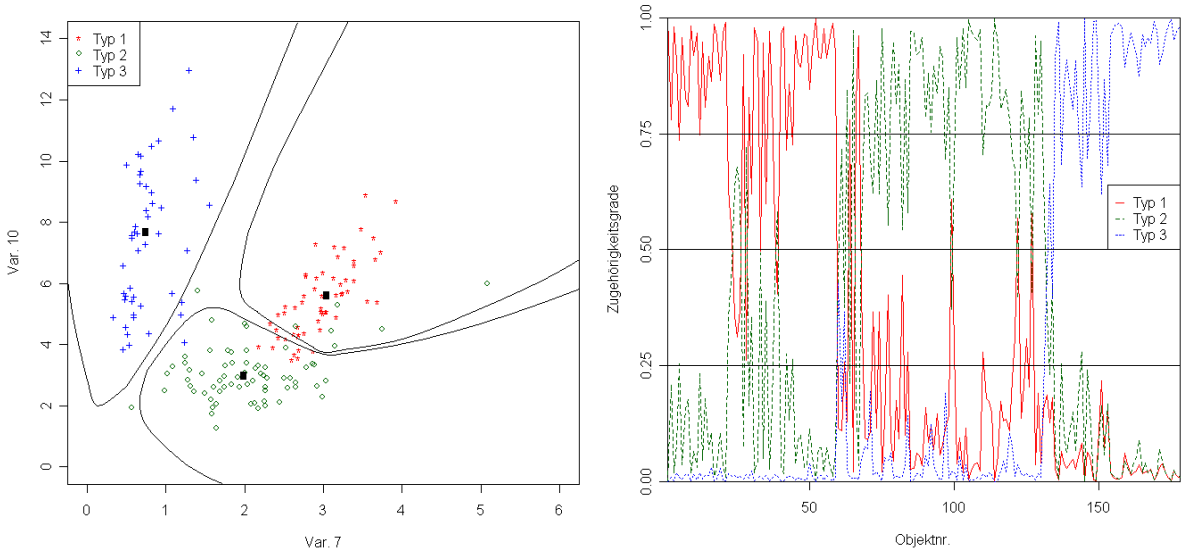
In den dargestellten Ergebnisgrafiken wird der einen Zugehörigkeitsgrad von mindestens 0.5 repräsentierende Bereich jeweils durch eine Ellipse bei den possibilistischen Ansätzen sowie durch eine Kurve bei der probabilistischen Analyse abgegrenzt. Eine zweite Grafik ermöglicht, die ermittelten Zugehörigkeitsgrade nachzuvollziehen. Der Weindatensatz besteht aus insgesamt 178 Instanzen, deren erste 59 sich Weintyp 1, die folgenden 71 Weintyp 2 und die letzten 48 Weintyp 3 zuordnen lassen; entsprechend sollten die jeweiligen Zugehörigkeitsgrade dominieren.

Aufgrund der elliptischen Form der Gruppen wurden die Gustafson-Kessel-Ansätze verwendet. Für den Ansatz zur Modellierung der Clusterabstoßung nach Timm wurden beide Erweiterungen näher betrachtet (Algorithmus 4.3 und Algorithmus 4.4), jeweils mit den Parametern $\gamma = 1$ und $\zeta = 1$. Für die Homogenitätsansätze wurden der Basisansatz (Algorithmus 4.5) und der Ansatz basierend auf der Dreiecksbeziehung der Distanzen (Algorithmus 4.7) mit den Parametern $\alpha = 0.5$, $\gamma = 1$ (Defaultwert) und $\zeta_i = \zeta \frac{n_i^\alpha \left(n_i^\alpha - 1\right)}{2n}$ mit $\zeta = 1$ (Defaultwert) einbezogen.

Die probabilstische Analyse liefert sinnvolle Clusterzentren, die sich augenscheinlich jeweils nahezu im Zentrum einer Weingruppe befinden (Abbildung 4.3a). Entsprechend kann auch in den Zugehörigkeitsgraden in Abbildung 4.3b eine deutliche Trennung zwischen den Clustern nachvollzogen werden, nur für die Objekte, die sich in der Überlappung der Gruppen befinden, ist die Trennung weniger offenkundig.

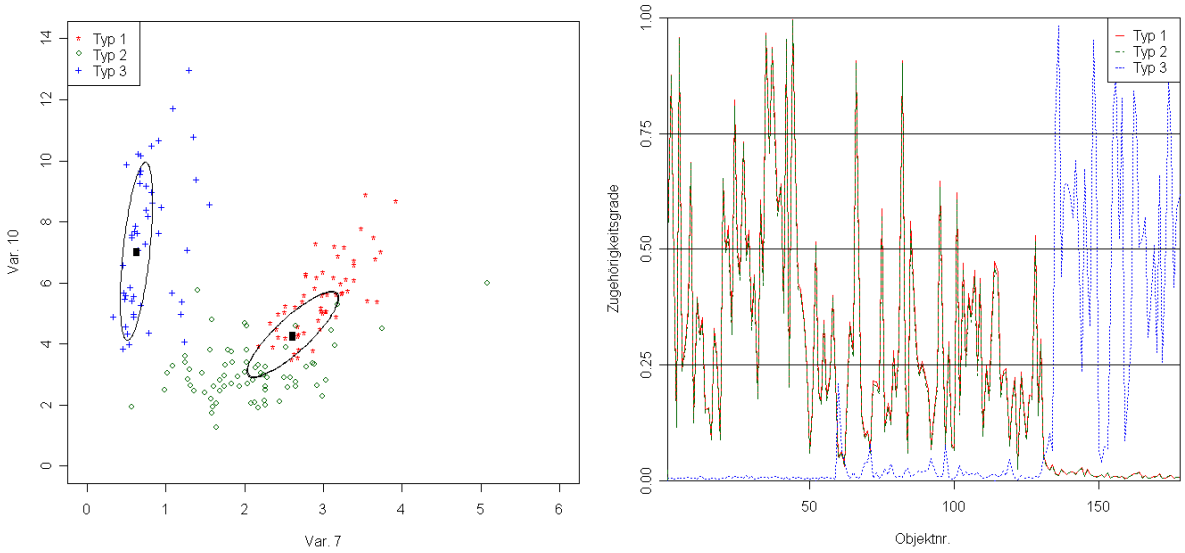
Der possibilistische Gustafson-Kessel-Algorithmus erkennt die gut separierte Gruppe der Weine von Typ 3, vermag jedoch nicht, die anderen Typen zu unterscheiden: Die Cluster erscheinen nahezu identisch (vgl. Abbildung 4.4a). Aufgrund der Clusteranziehung überlagern sich auch die Zugehörigkeitsgrade in Abbildung 4.4b für die Weine von Typ 1 und Typ 2 zu beiden Clustern.

[14] Der Weindatensatz besteht aus insgesamt 178 Weinen, die durch 13 quantitative Merkmale repräsentiert sind. Dabei lassen sich drei wesentliche Weinsorten unterscheiden. Aufgrund der vorhandenen Datenstruktur wird dieser Datensatz im Bereich des Data Minings bzw. des maschinellen Lernens häufig zum Testen von Klassifikations- und Clusteralgorithmen verwendet. Weiterführende Informationen sind unter Asuncion und Newman (2007) zu finden.



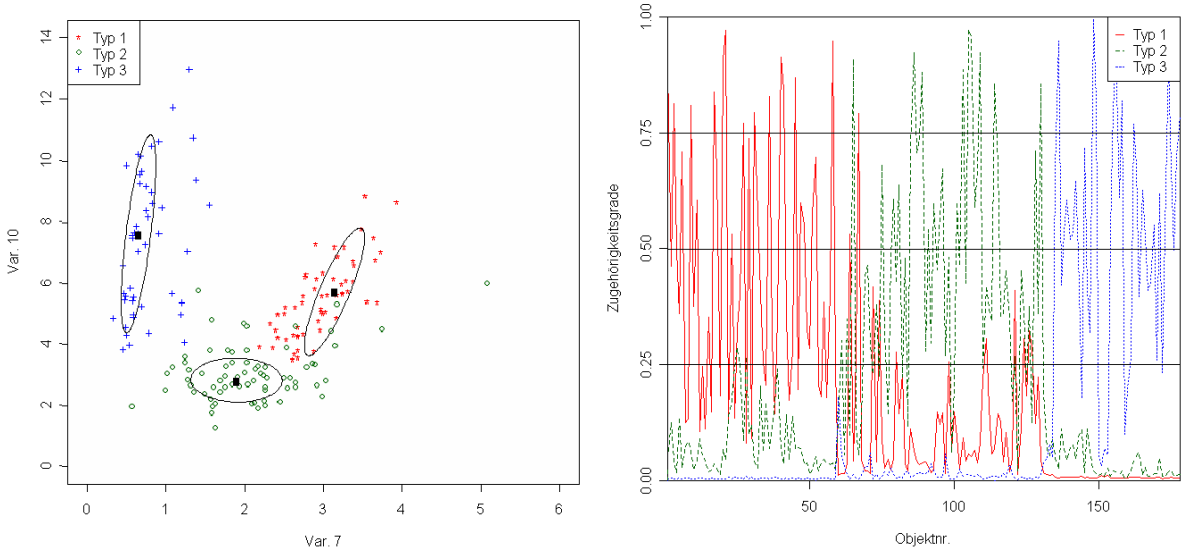
(a) Clusterprototypen (b) Zugehörigkeitsgrade

Abbildung 4.3.: Weindaten: Probabilistische GK-Analyse



(a) Clusterprototypen (b) Zugehörigkeitsgrade

Abbildung 4.4.: Weindaten: Possibilistische GK-Analyse



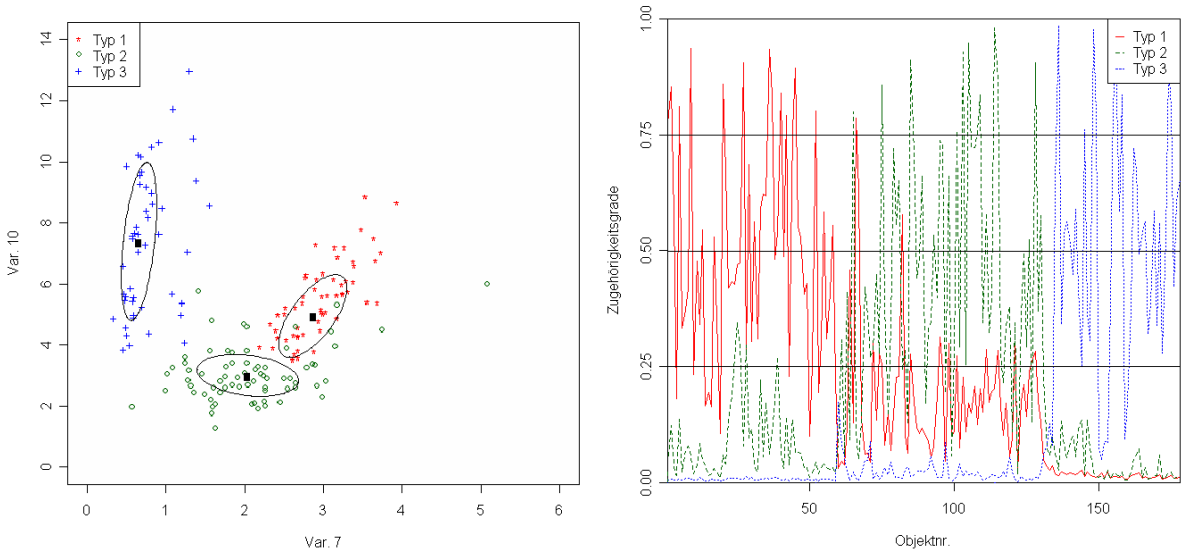
(a) Clusterprototypen (b) Zugehörigkeitsgrade

Abbildung 4.5.: Weindaten: Clusterabstoßung – Erweiterung 1

Die Ansätze basierend auf der Clusterabstoßung[15] und damit der Heterogenität zwischen den Clustern vermögen es, der Clusteranziehung entgegenzuwirken (vgl. Abbildung 4.5 bzw. Abbildung 4.6). Im Falle der stärkeren ersten Erweiterung (GK-Het1) erscheint jedoch in Abbildung 4.5a die Positionierung von zwei der drei Cluster weniger sinnvoll als die der schwächeren zweiten Erweiterung (GK-Het2) in Abbildung 4.6a. Die den Weintypen 1 und 2 zugehörigen Zentren werden durch die stärkere Abstoßung eher an den Rand der jeweiligen Cluster gedrängt; dies gilt insbesondere für das Zentrum von Typ 1. Aufgrund der Nichtberücksichtigung dichterer Regionen innerhalb der Cluster werden zudem relativ viele Objekte mit geringen Zugehörigkeitsgraden zu allen Clustern identifiziert, da sie zwischen diesen Clustern liegen (vgl. 4.5b). Dieser Problematik kommt bei der zweiten Erweiterung aufgrund des schwächeren Bestrafungsterms eine geringere Bedeutung zu, sie erzielt damit deutlich bessere Positionierungen der Clusterzentren. Es kann jedoch hier bereits festgestellt werden, dass die Ausrichtung des Clusters von Weintyp 2 aufgrund der Abstoßung von Typ 1 gegenüber der intuitiven Ausrichtung leicht rotiert zu sein scheint (vgl. Abbildung 4.6a).

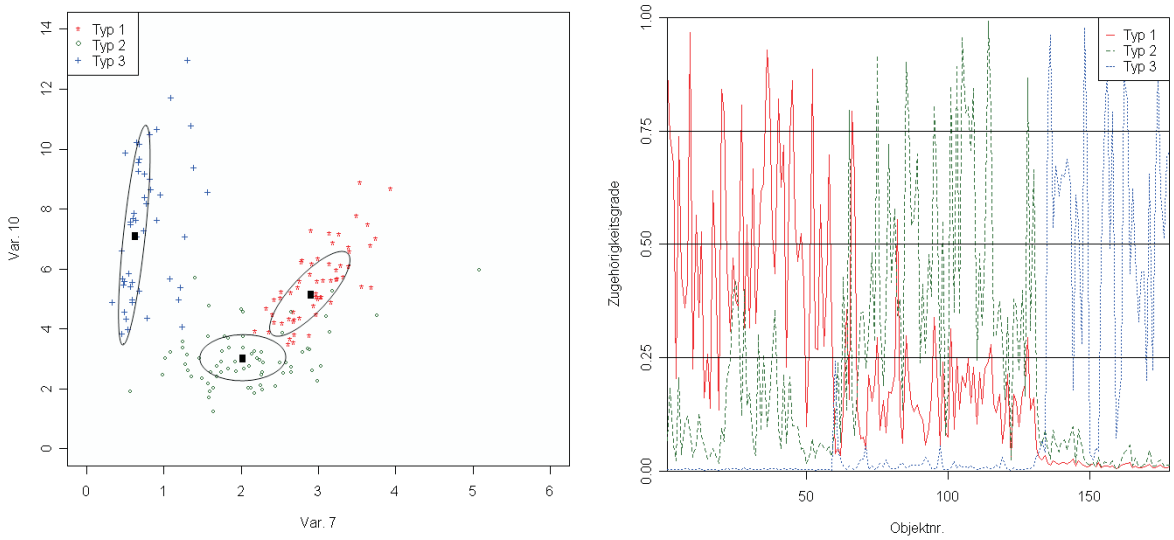
Die in Abschnitt 4.3 eingeführten Ansätze zur Modellierung der Clusterhomogenität ermöglichen es ebenfalls, alle drei Cluster zu unterscheiden (vgl. Abbildung 4.7 und Abbildung 4.8). Beide Ansätze führen zu intuitiven, vergleichbaren Clusterzentren, -formen und -ausrichtungen (vgl. Abbildung 4.7a und 4.8a). Auch die Zugehörigkeitsgrade zeigen, dass die Clusterprototypen eine relativ deutliche Clustertrennung anzeigen (vgl. Abbildung 4.7b und 4.8b). Die Struktur der Prototypen lässt sich durch die Wahl des Normierungsparameters ζ_i begründen: Durch die Stärkung des Bestrafungsterms im Vergleich zu den übrigen Termen der Zielfunktion erhält die Homogenitätsbedingung zusätzliches Gewicht, so dass sich die Clusterprototypen entsprechend der intuitiven Lage ausrichten.

[15] Die Ergebnisse unterscheiden sich von denen von Timm (2002, S. 64ff.) vorgestellten aufgrund der korrigierten Updateregeln (vgl. Abschnitt 4.2.2).



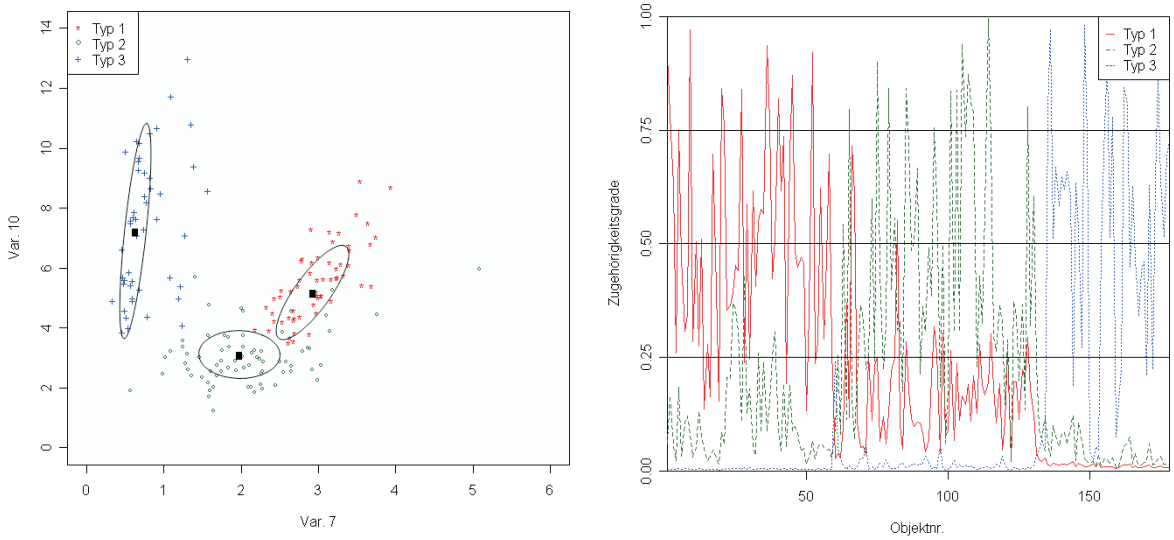
(a) Clusterprototypen (b) Zugehörigkeitsgrade

Abbildung 4.6.: Weindaten: Clusterabstoßung – Erweiterung 2



(a) Clusterprototypen (b) Zugehörigkeitsgrade

Abbildung 4.7.: Weindaten: Clusterhomogenität – Basisansatz



(a) Clusterprototypen (b) Zugehörigkeitsgrade

Abbildung 4.8.: Weindaten: Clusterhomogenität – Dreiecksbeziehung der Distanzen

4.4.2. Vergleich der Validitätsmaße

Neben dem Vergleich der Clusterausrichtungen und der Zugehörigkeitsgrade wurden die Algorithmen mit Hilfe verschiedener Validitätsmaße gegenübergestellt (vgl. Abschnitt 3.5); es wurden der Separationsindex (3.15) nach Xie und Beni, die Partitionsentropie (3.14) sowie die mittlere Partitionsdichte (3.16) verwendet. Für die in Abschnitt 4.3 vorgestellten, auf der Clusterhomogenität basierenden Ansätze wurden die Validitätsmaße für alternative α-Werte bestimmt und so ein sinnvoller Bereich für α ermittelt. Die folgenden Grafiken enthalten nur Angaben für $\alpha \in [0.2, 0.8]$, da für $\alpha < 0.2$ nahezu alle Objekte in den Bestrafungsterm einbezogen werden. Dies steht in direktem Widerspruch zur grundsätzlichen Idee dieses Ansatzes. Für $\alpha > 0.8$ werden Ergebnisse erzielt, die sich, je höher α gewählt wird, immer mehr der klassischen possibilistischen Analyse nach Krishnapuram und Keller (1993) angleichen. Bei $\alpha = 1$ sind die Ergebnisse zumeist identisch; der Fall eines Objekts mit einem Zugehörigkeitsgrad von Eins, d.h. eine Überlagerung von Clusterzentrum und Objekt, tritt vernachlässigbar selten auf.

Die übrigen Ansätze, d.h. die probabilistische sowie die klassische possibilistische Analyse und die Ansätze basierend auf der Clusterabstoßung, sind unabhängig von α und darum in den Grafiken als Konstanten angegeben. Die Anführung der Validitätsmaße der probabilistischen Analyse geschieht der Vollständigkeit halber, da sie aufgrund der abweichenden Interpretation der Zugehörigkeitsgrade und der diese betreffenden zusätzlichen Bedingung nur bedingt vergleichbar sind.

Separationsindex

Der Separationsindex nach Xie und Beni wurde angewandt, da er die grundsätzliche Idee des Ansatzes von Timm (2002) aufnimmt, indem er die Distanz zwischen den Clusterzentren einbezieht und damit eine deutlichere Clustertrennung bevorzugt (vgl. Abschnitt 4.2). Die Ergebnisse

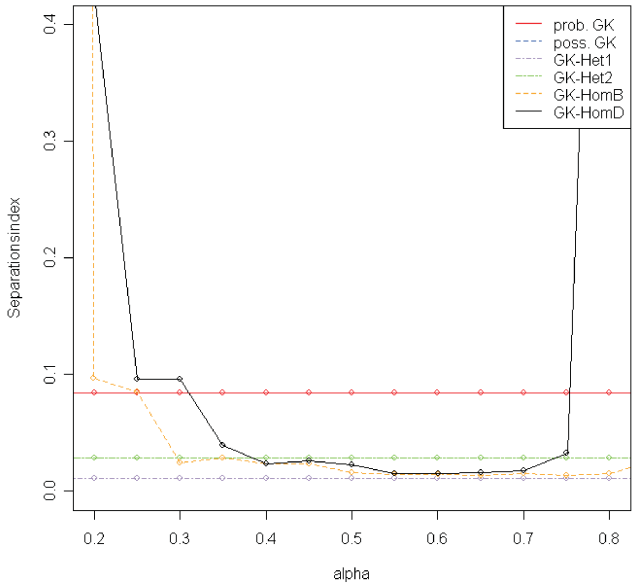


Abbildung 4.9.: Separationsindizes der einzelnen Verfahren

der probabilistischen Analyse sind aufgrund der Bedingung in (3.3) im Allgemeinen schlechter als diejenigen, die durch eine possibilistische Analyse erzielt werden. Unter Beachtung des Separationsindexes erzielt die erste Erweiterung basierend auf der Clusterheterogenität aufgrund der starken Clusterabstoßung das beste Ergebnis (GK-Het1 in Abbildung 4.9). Dies entspricht der Idee dieses Ansatzes, der gezielt auf die deutliche Clustertrennung ausgerichtet ist. Der zweite, schwächere Ansatz zur Heterogenität (GK-Het2) schließt durch die entsprechende Schwächung des Bestrafungsterms bereits mit schlechterem Ergebnis ab. Für einen Wert $\alpha \in [0.4, 0.75]$ erzielen sowohl der Basisansatz unter Beachtung der Clusterhomogenität (GK-HomB) als auch der Ansatz unter Berücksichtigung der Dreiecksbeziehung der Distanzen (GK-HomD) vergleichbare Ergebnisse mit den Ansätzen von Timm (2002); teilweise ist die Trennung sogar besser als die durch die schwächere Clusterabstoßung erlangte (GK-Het2).

Der Separationsindex des originalen possibilistischen Ansatzes nach Krishnapuram und Keller (1993) wird in Abbildung 4.9 nicht angegeben, da er aufgrund der mangelnden Trennung der Weintypen 1 und 2 mit einem Separationsindex von $S(U) = 744.517$ zu hoch ist, um in der Grafik dargestellt zu werden, da aufgrund der Skalierung sonst keine Unterscheidung der übrigen Ergebnisse mehr möglich ist (vgl. Abbildung 4.4a).

Partitionsentropie

Um den erhaltenen Informationsgehalt zu evaluieren, erzielt die probabilistische Clusteranalyse unter Beachtung der Partitionsentropie i.d.R. als Folge der wahrscheinlichkeitsorientierten Zugehörigkeitsgrade und der daraus resultierenden meist härteren Zuordnung bessere Ergebnisse als possibilistische Ansätze. Der erste Ansatz zur Clusterabstoßung (GK-Het1) liefert auch hier das beste Ergebnis der possibilistischen Analysen (Abbildung 4.10); dies lässt sich durch die geringen Zugehörigkeitsgrade bei einer hohen Anzahl der Objekte begründen (vgl. Abbildung 4.5b). Die übrigen possibilistischen Ansätze erreichen insgesamt für $\alpha \in [0.4, 0.75]$ ähnlich gute

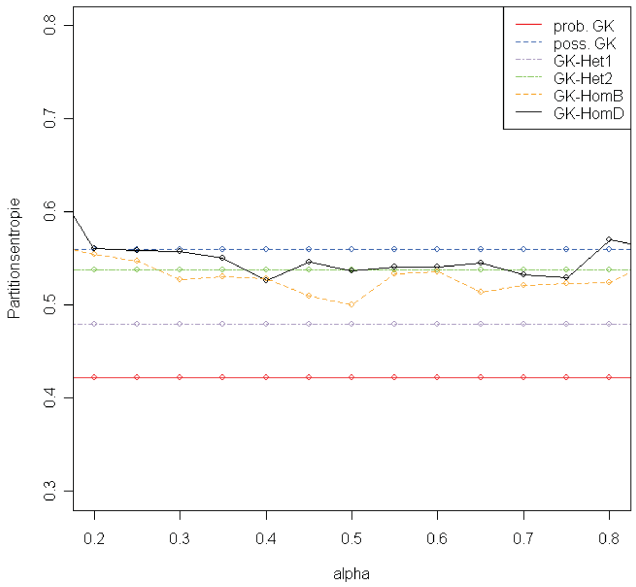


Abbildung 4.10.: Partitionsentropien der einzelnen Verfahren

Ergebnisse. Dabei muss die Partitionsentropie der klassischen possibilistischen Analyse mit Vorsicht betrachten werden, da in diesem Fall nur zwei Cluster durch den Algorithmus identifiziert werden können.

Mittlere Partitionsdichte

Die mittlere Paritionsdichte bewertet die Homogenität der einzelnen Cluster und entspricht damit der Grundidee der darauf basierenden Ansätze aus Abschnitt 4.3. Zur Berechnung der mittleren Partitionsdichte nach (3.16) werden die clusterspezifischen Kovarianzmatrizen Σ_i benötigt. Da diese bei den Erweiterungen zur Clusterabstoßung und zur Clusterhomogenität jedoch nicht explizit bestimmt werden, sondern lediglich jeweils eine Annäherung ihres Vielfachen S_i sowie die positiv definite Normmatrix A_i bekannt sind, wurden sie zur Berechnung der mittleren Partitionsdichte basierend auf den Clusterergebnissen mit (3.8) gesondert ermittelt.

Die probabilistische Analyse führt in den meisten Fällen zu schlechteren Ergebnissen als die possibilistischen Ansätze, da alle Objekte einen verhältnismäßig starken Einfluss auf die Positionierung der Clusterprototypen ausüben. Die Ansätze basierend auf der Clusterabstoßung erzielen beide im Mittel eine höhere Dichte innerhalb der Cluster (Abbildung 4.11). Die Ansätze zur Clusterhomogenität konzentrieren sich auf eine hohe Ähnlichkeit innerhalb der ermittelten Cluster. Bei Wahl geeigneter α-Werte ist es daher möglich, eine sehr hohe mittlere Partitionsdichte zu erzielen. Mit Werten für $\alpha \in [0.45, .075]$ erzielen diese Ansätze daher die besten Ergebnisse. Die mittlere Partitionsdichte der klassischen possibilistischen Analyse erscheint auch hier nicht wesentlich schlechter als die der übrigen possibilistischen Verfahren; dies verdeutlicht, dass die isolierte Betrachtung einzelner Validitätsmaße nicht ausreicht.

Aufgrund der hier vorgestellten sowie weiterer Ergebnisse konnte empirisch $\alpha \in [0.4, 0.6]$ als anzuwendendes Intervall ermittelt werden (vgl. Neumann, 2008, S. 60ff.). Damit werden genü-

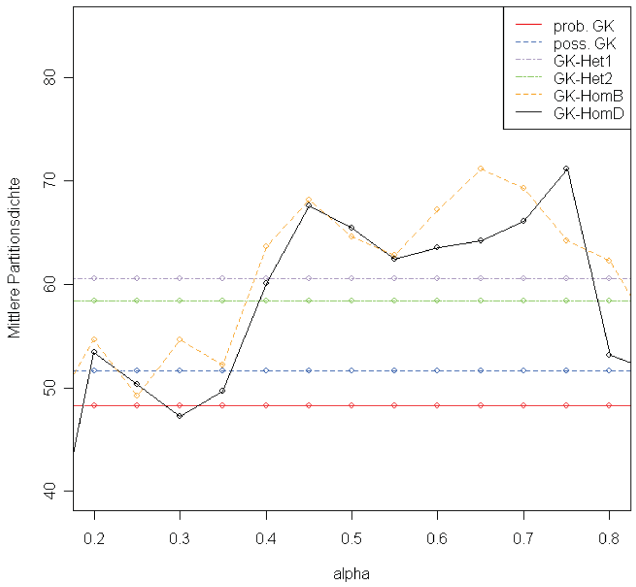


Abbildung 4.11.: Mittlere Partitionsdichten der einzelnen Verfahren

gend Objekte durch den Bestrafungsterm berücksichtigt, ohne dass der grundsätzlichen Idee der Clusterhomogenität entgegengewirkt wird. Ferner ist eine Stärkung des Bestrafungsterms durch eine geeignete Wahl des Normierungsparameters ζ_i sinnvoll, da, je stärker der Bestrafungsterm im Vergleich zu den übrigen Termen der Zielfunktion ist, desto weniger anfällig das Verfahren gegenüber der Clusteranziehung ist, weil der Struktur des einzelnen Clusters eine höhere Bedeutung zukommt.

Weitere Analysen und Vergleiche wurden auf dem Iris-Datensatz (Asuncion und Newman, 2007) sowie auf verschiedenen künstlich erzeugten Datensätzen durchgeführt (Neumann, 2008). Im Rahmen dieser Analysen wurden weitestgehend vergleichbare Ergebnisse bzgl. der Eignung der einzelnen Algorithmen erzielt.

Mehr als die Vergangenheit interessiert mich die Zukunft, denn in ihr gedenke ich zu leben.

A. Einstein

5 Dynamisches Fuzzy-Clustering

In der Literatur gibt es verschiedene Ansätze zur dynamischen Untersuchung von Clusterstrukturen. Daher wird vor der eigentlichen Untersuchung der grundlegenden Fragestellung zur Analyse sich verändernder Strukturen in Abschnitt 5.1 eine Einführung in den Themenbereich gegeben, in deren Zusammenhang der Begriff der dynamischen Analyse im vorliegenden Kontext erläutert wird und verschiedene verwandte Forschungsgebiete aufgezeigt werden. Im Anschluss werden in Abschnitt 5.2 die grundlegenden Aspekte der dynamischen Untersuchung im relevanten Kontext eingeführt, bevor in den Abschnitten 5.3 bis 5.7 schließlich auf die möglichen Veränderungen in der Clusterstruktur im Detail eingegangen wird. Zusammenfassend wird in Abschnitt 5.8 eine Übersicht über die einzelnen Analyseschritte gegeben.

5.1. Einführung und Related Work

Der Begriff der *dynamischen* Clusteranalyse wird für unterschiedliche Bereiche der Clusteranalyse bei zeitabhängigen Untersuchungen verwendet. Dabei mangelt es häufig an einer expliziten Unterscheidung, welcher Teilbereich einer Problemstellung als dynamisch zu betrachten ist: Zum einen können die untersuchungsrelevanten Objekte dynamisch sein, d.h., es handelt sich um kontinuierliche Datenströme bzw. um die Untersuchung von Zeitreihen o.ä., zum anderen kann auch eine dynamische Clusterstruktur analysiert werden, bei der die Veränderungen innerhalb dieser Struktur im Sinne des Change Minings aufgedeckt werden. Weber (2007, S. 317) hat hierfür die in Tabelle 5.1 dargestellte Kategorisierung nach Art der Dynamik vorgenommen.

	Statische Cluster	Dynamische Cluster
Statische Objekte	klassische Clusteranalyse	Clustering einer Menge sich verändernder Eigenschaftsvektoren
Dynamische Objekte	Clustering von zeitabhängigen Daten	Clustering einer Menge sich verändernder zeitabhängiger Daten

Tabelle 5.1.: Kategorisierung von Clusterproblemen

Handelt es sich sowohl um statische Objekte, die durch ihre Eigenschaftsvektoren repräsentiert werden, als auch um eine statische Clusterstruktur, so findet eine rein statische Clusteranalyse Anwendung; das zugehörige klassische Analysevorgehen wird in Kapitel 3 grundlegend beschrieben und ist darüber hinaus für eine dynamische Untersuchung nicht von Bedeutung. Die anderen drei Felder in Tabelle 5.1 stehen für die unterschiedlichen Arten der dynamischen Clusteranalyse.

Soll innerhalb von zeitabhängigen Daten eine statische Clusterstruktur ermittelt werden, so erfolgt die Anwendung statischer Clusteranalyseverfahren für zeitabhängige Daten wie Trajektorien im Eigenschaftsraum oder auch kontinuierliche Datenströme wie Funktionen oder Zeitreihen, die einzelnen unveränderlichen Clustern zugewiesen werden. Diese Art der Clusteranalyse wird auch als temporale Clusteranalyse bezeichnet (vgl. u.a. Gudmundsson u. a., 2008; Kalnis u. a., 2005)[16]. Ein Beispiel hierfür stellt das einfache Clustering von Clickstreams dar. In Online-Shops ist es auf diese Weise möglich, Navigationen auf den Seiten eines Shops nachzuvollziehen und zu gruppieren, so dass gängige Nutzerpfade herausgefiltert werden können. Zur temporalen Analyse zeitabhängiger Daten wird in der Literatur eine Vielzahl von Verfahren vorgeschlagen (vgl. u.a. Angers, 2002; Tan u. a., 2013). So führen z.B. Joentgen u. a. (1999) als möglichen Algorithmus den *Functional Fuzzy-C-Means* (FFCM) ein, eine Erweiterung des klassischen Fuzzy-C-Means zur Anwendung auf Funktionenverläufe und Szenarien. Angers (2002) verwendet zur Zeitreihenanalyse ein Verfahren, das auf der Ermittlung von Ähnlichkeiten zwischen Wavelets beruht. Zur Clickstreamanalyse und dem Clustern von Navigationspfaden auf Internetseiten entwickelten Ali und Ketchpel (2003) den *Golden Path Analyzer*, der den Anwender dabei unterstützen soll, häufige Klickwege einer Webseite aufzudecken und darauf aufbauend die Navigation zu erleichtern. Auf diese Weise ließe sich eine Vielzahl möglicher Anwendungen und Algorithmen aufzählen, da dieses Teilgebiet der dynamischen Analyse bereits Teil von intensiven Untersuchungen geworden ist.

Sind die betrachteten Objekte durch ihre Eigenschaftsvektoren gegeben, die in dem jeweils betrachteten Zeitfenster in statischer Form vorliegen, sich jedoch zwischen den einzelnen Zeitfenstern verändern können, kann dies eine Veränderung der Clusterstruktur nachsichziehen. In diesem Zusammenhang erfolgt eine Analyse der im jeweiligen Zeitfenster statischen Objekte, um die Veränderungen der Clusterstruktur aufzudecken, d.h., die Cluster sind als dynamisch anzusehen. Bei dieser Form der dynamischen Clusteranalyse handelt es sich um eine Analyse im Sinne des Change Minings (vgl. Abschnitt 1.3), da das Ziel im Aufdecken der Veränderungen innerhalb der Clusterstruktur besteht. Diese Analyse ist besonders im Bereich des Marketings von großer Bedeutung (vgl. Abschnitt 1.4). So weisen die Kunden eines Unternehmens innerhalb eines Zeitfensters ein statisches Kaufverhalten auf, die Struktur der Kundensegmente verändert sich jedoch i.d.R. im Laufe der Zeit. Ist im Rahmen der dynamischen Analyse die Untersuchungsgruppe fixiert, z.B. durch Verwendung eines Panels, kann dennoch die Entwicklung der einzelnen Kunden zwischen den jeweiligen Untersuchungszeitpunkten nachvollzogen werden; lediglich für die Analyse selbst wird nur eine statische Momentaufnahme herangezogen. Alternativ kann auch eine sich ständig verändernde, anonyme Untersuchungsgruppe wie z.B. die Käuferschaft eines Supermarktes analysiert werden, ohne dass die Analyse der Kundengruppen negativ beeinflusst wird. Aufgrund der Relevanz der sich verändernden Clusterstrukturen im Marketingkontext steht diese Kategorie der dynamischen Clusteranalyseprobleme im Sinne

[16] Die temporale Clusteranalyse gehört zu den Verfahren des temporalen Data Minings (engl.: *Temporal Data Mining*). Antunes und Oliveira (2001) geben eine Einführung in dieses Teilgebiet des Data Minings.

des Change Minings in dieser Arbeit im Fokus und wird in den folgenden Abschnitten näher untersucht. Ausgangspunkt sind dabei Eigenschaftsvektoren, auf deren Basis die dynamische Entwicklung von Clustern analysiert werden soll.

Die temporale Clusteranalyse kann ebenfalls im Rahmen des Change Minings durchgeführt werden, wenn sowohl die Objekte in dynamischer Form durch ihre Trajektorien gegeben sind als auch eine sich verändernde Clusterstruktur analysiert werden soll, so dass beide Seiten der Untersuchung, d.h. Input und Output, als dynamisch anzusehen sind. In diesem Fall wird häufig keine Clusteranalyse des gesamten Datenstroms vorgenommen, sondern Teilströme werden analysiert, die in den betrachteten Zeitfenstern aufgetreten sind (vgl. Aggarwal u. a., 2003). Auf diese Weise können Veränderungen in der Clusterstruktur der Datenströme festgehalten werden. Zur Analyse von Veränderungen in der Clusterstruktur von Datenströmen wurden in den letzten Jahren zahlreiche Untersuchungen angestellt. Dong u. a. (2003) geben eine grundlegende Einführung in diesen Themenkomplex und schlagen grundsätzliche Vorgehensweisen vor, die von anderen Autoren in verschiedenen Veröffentlichungen konkretisiert werden. So führen Nasraoui u. a. (2003) einen Ansatz zum inkrementellen Lernen bei der Analyse von Clickstreams ein. Cao u. a. (2006) und Zhou u. a. (2008) untersuchen die generelle Entwicklung in der Clusterstruktur der Datenströme, wobei letztere insbesondere auf die Ermittlung einer geeigneten Clusterzahl innerhalb der einzelnen Zeitintervalle fokussieren. Angstenberger (2000) widmete im Rahmen ihrer Doktorarbeit dieser Art der dynamischen Betrachtung der Clusterstruktur von Datenströmen eine ausführliche Untersuchung. Einige Aspekte der dynamischen Clusteranalyse von Zeitreihen und Datenströmen sind auch auf die dynamische Analyse statischer Momentaufnahmen übertragbar, wie in den folgenden Abschnitten verdeutlicht wird.

5.2. Generelle Aspekte

Zum Change Mining im Rahmen der Clusteranalyse erscheint das Fuzzy-Clustering wesentlich sinnvoller als harte Analysemethoden, da zwar eine Clusterhierarchie mit Hilfe inkrementellen Lernens dynamisch angepasst werden kann (vgl. Nassar u. a., 2004), bei partitionierenden Verfahren und der Repräsentation von Clustern mittels Clusterprototypen jedoch andernfalls eine leichte Veränderung einzelner Objekte und dem damit einhergehenden potentiellen Wechsel des zugehörigen Clusters die Clusterprototypen aufgrund der harten Zugehörigkeitsgrade stark beeinflusst werden können (vgl. Abschnitt 3.4 sowie Kapitel 4). Um die durch die untersuchten Objekte hervorgerufenen tatsächlichen, meist graduellen Veränderungen nachvollziehen und die daraus resultierenden Änderungen der Clusterstruktur identifizieren zu können, stellen die kontinuierlichen Zugehörigkeitsgrade des Fuzzy-Clusterings ein elementares Werkzeug dar (vgl. Crespo und Weber, 2005). Crespo und Weber (2005) verwenden hierbei probabilistische Verfahren zur Clusteranalyse (Abschnitt 3.3). Diese erweisen sich jedoch aufgrund ihrer wahrscheinlichkeitstheoretischen Grundlage als eher ungeeignet, um verschiedene abrupte Veränderungen innerhalb der Clusterstruktur aufzudecken. Da der Fokus bei Crespo und Weber (2005) jedoch ausschließlich auf der Bestimmung einer neuen Clusterzahl auf Basis der aktuellen Clusterstruktur liegt, die tatsächlichen Änderungen innerhalb der Struktur jedoch nicht näher untersucht werden, ist dieses Vorgehen ausreichend. Wird hingegen anstelle der probabilistischen eine possibilistische Clusteranalyse zugrundegelegt, so wird der Einfluss einzelner Objektveränderungen aufgrund der Bestimmung der Clusterprototypen unabhängig von der La-

ge der übrigen Cluster verringert (Abschnitt 3.4). Des Weiteren ermöglicht die possibilistische Analyse unter anderem das Erkennen von Ausreißern. Diese dürfen im Change Mining nicht vernachlässigt werden, da sie unter anderem zukünftig neu entstehende Cluster repräsentieren können (vgl. Abschnitt 5.4).

Ziel der dynamischen Analyse einer Clusterstruktur im Sinne des Change Minings ist das Erkennen von Veränderungen innerhalb dieser Struktur, um darauf aufbauend zukünftig zu erwartende Entwicklungen prognostizieren zu können; der Fokus liegt also neben ausschließlich deskriptiven Modellen wie den von Angstenberger (2000) und Crespo und Weber (2005) eingeführten ebenso auf der Vorhersage zukünftiger Entwicklung, d.h., die spezifizierten Modelle weisen außerdem einen prädiktiven Charakter auf. Dabei ist die Vorhersage der reinen Beschreibung übergeordnet, da außer den deskriptiven weitere Eigenschaften im Verlauf der Analyse spezifiziert werden müssen (vgl. Böttcher u. a., 2008). Im Rahmen des Marketings ist eine solche Analyse von hoher Bedeutung: Werden z.B. Veränderungen in der Kundenstruktur erkannt, kann ein Unternehmen darauf reagieren und seine Marketingmaßnahmen gezielt daran anpassen. Verfügt es zusätzlich über Anhaltspunkte, wie sich der Markt in naher Zukunft entwickeln wird, so vermag es außerdem, bereits vorab Maßnahmen einzuleiten, um allgemeinen Entwicklungen zu begegnen und unerwünschten, soweit möglich, entgegenzuwirken.

Generell wird zwischen abrupten und graduellen Veränderungen innerhalb einer Clusterstruktur unterschieden. Abrupte Veränderungen sind Abweichungen der gesamten Clusterstruktur von einem bisher bekannten Muster, die i.d.R. eine direkte Reaktion erfordern; sie gehen im Allgemeinen mit einer Veränderung der Clusterstruktur bzgl. der Clusterzahl einher. Folgende vier abrupte Veränderungen in der Clusterstruktur müssen bei der dynamischen Analyse berücksichtigt werden (vgl. u.a. Crespo und Weber, 2005; Zhou u. a., 2008):

1. Neubildung von Clustern
2. Eliminierung von Clustern
3. Vereinigung mehrerer zu einem gemeinsamen Cluster
4. Teilung eines Clusters in mehrere separate Cluster

Die graduellen Veränderungen beinhalten Abweichungen einzelner Clusterprototypen unter Beibehaltung der generellen Aufteilung. Diese Veränderungen können neben der Darstellung der sich ändernden Prototypen als Indikator für zukünftige abrupte Veränderungen dienen. So kann sich die Position eines Clusters verändern, d.h., das Clusterzentrum weist eine Verschiebung bzgl. der untersuchten Eigenschaften auf; weiterhin kann die Dichte und die Verteilung innerhalb eines Clusters variieren. Nimmt z.B. die Dichte eines Clusters stetig ab, impliziert dies eine anstehende Clustereliminiation, da nicht ausreichend neue Objekte diesem Cluster zugeordnet werden können. Außerdem lassen sich die Entwicklungen der Cluster über mehrere Perioden hinweg durch Clustertrajektorien abbilden, auf deren Basis ihre räumliche Veränderung nachvollzogen werden kann. Auf diese Weise kann z.B. ein mögliches Annähern von zwei Clustern und eine dadurch angedeutete potentielle Vereinigung dieser frühzeitig erkannt werden. Tabelle 5.2 zeigt verschiedene abrupte Veränderungen, die durch wiederholt auftretende graduelle Veränderungen impliziert werden können.

Aus dem Zusammenhang zwischen den graduellen und den abrupten Veränderungen lässt sich folgern, dass nicht nur die Veränderung der Clusterstruktur, sondern auch die Entwicklung innerhalb eines Clusters von Bedeutung ist (vgl. Zhou u. a., 2008). Besonders elementar ist die

Graduelle Veränderung bzgl.	Mögliche abrupte Veränderung
Clusterposition	Vereinigung von Clustern
Clusterdichte	Clusterelimination; Trennung eines Clusters
Clustervolumen	Clusterelimination; Trennung eines Clusters

Tabelle 5.2.: Zusammenhang gradueller und abrupter Veränderungen

Analyse aufgrund der Tatsache, dass die Objektmenge innerhalb der Cluster nicht während der gesamten Clusterlebenszeit dieselbe ist; Objekte können zu einem Cluster hinzustoßen oder ein Cluster verlassen (vgl. Kalnis u. a., 2005). Außerdem ist nicht gewährleistet, dass die untersuchten Objekte zu jedem Analysezeitpunkt dieselben sind. Gerade im Marketing und in der Marktforschung ist es durchaus möglich, dass eine untersuchte Kundengruppe von der in der Folgeperiode analysierten abweicht. Eine weitestgehend gleichbleibende Stichprobe ist nur dann möglich, wenn die Auswahl mit Hilfe von Panels erfolgt. Soll jedoch z.B. das Kaufverhalten eines Supermarktes analysiert werden, so sind die einzelnen Kunden nicht länger identifizierbar, d.h., die Untersuchungsgruppe variiert ggf. stark. In diesem Fall muss auf eine Objektidentifikation vollständig verzichtet werden; es erfolgt lediglich eine Analyse der allgemeinen Clusterstruktur. Dieses Vorgehen soll im Folgenden Anwendung finden.

Für das Change Mining bedeutet der Zusammenhang zwischen graduellen und abrupten Veränderungen sowie der Fokus auf die allgemeine Clusterstruktur, dass zusätzlich zu den allgemeinen Clustereigenschaften weitere Informationen in einer Historie abgelegt werden müssen, um daraus die Veränderung der Charakteristika ableiten zu können. Zhou u. a. (2008) bezeichnen dieses Vorgehen als *in-cluster maintenance*, d.h., es werden die benötigten Clusterinformationen beibehalten, um die fortlaufende Entwicklung des Clusters nachvollziehen zu können.

Unabhängig von der Art der Veränderung ist das Vorhandensein von Hintergrundwissen aus den vorangegangenen Perioden elementar, um die Anpassungen der Clusterstruktur messen und nutzbar machen zu können; dementsprechend ist dieses Wissen ein natürlicher Teil der Problemdefinition, wenn man sich mit dem Change Mining in einer Clusterstruktur befasst (vgl. Pechoucek u. a., 1999). Bevor die Veränderungen jedoch untersucht werden können, müssen die zeitlichen Parameter festgelegt werden, in denen eine Analyse vorgenommen wird. Hierzu werden Zeitintervalle definiert, in denen eine statische Momentaufnahme der Situation anhand entsprechender Periodendaten erfolgt (vgl. z.B. Apeh und Gabrys, 2011; Crespo und Weber, 2005). Die Länge eines solchen Intervalls zwischen dem Erstellen einer Clusterstruktur und ihrem Update ist anwendungsabhängig, da in verschiedenen Szenarien Veränderungen unterschiedlich schnell auftreten und analysiert werden können (vgl. Abschnitt 1.3). Für eine Analyse sind jeweils nur die aktuellen Objekteintragungen relevant, vorangegangene sollen keinen oder nur geringen Einfluss auf die Clusteranpassung erhalten. Diese Forderung kann unter Anwendung von sich überlappenden Zeitfenstern erfolgen; dabei werden entweder alle Daten des Zeitintervalls gleichermaßen einbezogen oder es erfolgt eine Gewichtung auf Basis ihres Alters. Bei der ersten Möglichkeit wird vorab die Länge τ eines Zeitfensters festgelegt, aus dem alle Objektdaten in die Analyse einbezogen werden. Ferner wird ein Parameter Δt gewählt, durch den die zeitliche Veränderung des Zeitfensters und damit der jeweilige Zeitpunkt der

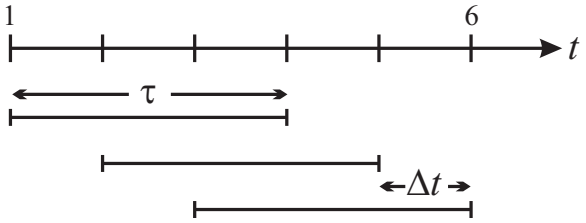


Abbildung 5.1.: Überlappende Zeitfenster

Analyse bestimmt wird. I.d.R. gilt $\Delta t \leq \tau$. Abbildung 5.1 soll dieses Vorgehen verdeutlichen: Sind die Daten z.B. in Form von Monatsdaten gegeben, so können jeweils zum Monatsende die Daten der letzten drei Monate analysiert werden; in diesem Fall ist $\tau = 3$ und $\Delta t = 1$. Dies bedeutet, dass nach der Analyse potentiell auftretender Unterschiede die neue Clusterstruktur anhand der Daten der letzten drei Monate erlernt wird, falls entsprechende Veränderungen zu verzeichnen sind. Wird $\tau = \Delta t$ gewählt, so ist die Überlappung der Zeitfenster ausgeschlossen und jedes Zeitfenster wird separat zum Update der Struktur hinzugezogen.

Werden im Gegensatz zu der zuvor beschriebenen Vorgehensweise die Objektdaten abhängig von ihrem Alter gewichtet, so erhalten diejenigen Daten mit einem geringeren Grad an Aktualität ein kleineres Gewicht (vgl. Aggarwal u. a., 2003; Böttcher u. a., 2008). Auf diese Weise gehen nicht alle Daten unabhängig vom Alter innerhalb des Zeitfensters gleichermaßen in die Untersuchung ein. Aggarwal u. a. (2003) schlagen vor, ein pyramidenartiges Zeitfenster zu wählen, so dass aus älteren Teilintervallen weniger Daten eingehen als aus neueren. Dieses Vorgehen hat jedoch den Nachteil, dass die Repräsentativität der ausgewählten Daten aus den älteren Teilintervallen gewährleistet sein muss, damit das Ergebnis nicht verfälscht wird. Im Fall der von Böttcher u. a. (2008) vorgeschlagenen Alterungsfunktion (vgl. Abschnitt 1.3) stellt sich diese Problematik nicht, da lediglich die Gewichtung der einzelnen Objektdaten variiert. Eine separate Festlegung der Länge des Zeitfensters ist dabei nicht zwingend erforderlich, da die Nichtberücksichtigung zu alter Daten mit Hilfe der Alterungsfunktion modelliert werden kann.

Sind die zeitlichen Parameter bzgl. Länge des Zeitintervalls und der Einbeziehung weniger aktueller Daten festgelegt, kann die Analyse der Veränderungen erfolgen. Hierzu kann ausgehend von der bisherigen Struktur zunächst ein Reclustering der Objekte aus dem aktuellen Zeitintervall erfolgen; die zuvor ermittelten Prototypen dienen der Initialisierung der erneuten Clusteranalyse. Auf diese Weise sind genaue Untersuchungen möglich, da die aktuelle Struktur bei gleicher Clusterzahl vollständig erkannt wird. Aufgrund des Aufwands einer vollständigen Clusteranalyse ist dieses Vorgehen jedoch nur bei kleinen Datensätzen empfehlenswert, da im Anschluss eine Zuordnung der neuen zu den alten Clustern erfolgen muss, so dass die tatsächlichen Veränderungen untersucht werden können. Diese Untersuchung ist sehr komplex, insbesondere wenn es darum geht, einfache Schwankungen von relevanten Veränderungen innerhalb der Struktur zu differenzieren. Ferner muss am Ende der Veränderungsanalyse eine weitere Clusteranalyse mit der angepassten Clusterzahl durchgeführt werden, sofern abrupte Veränderungen aufgetreten sind; hierdurch wird der Aufwand noch einmal zusätzlich erhöht. Aufgrund der erhöhten Komplexität und der nicht vorab sinnvoll bestimmbaren relevanten Clusterzahl wird daher anstelle eines Reclusterings eine Zuordnung der für die aktuelle Periode

relevanten Objekte[17] gemäß der bekannten Clusterstruktur vorgenommen. Dazu erfolgt eine Bestimmung der jeweiligen Distanzen zu den Clusterzentren sowie die Bestimmung der entsprechenden Zugehörigkeitsgrade nach (3.12), um aufgetretene Veränderungen darstellen und evaluieren zu können. So lässt sich bereits vor der Durchführung einer erneuten Clusteranalyse am Ende der Veränderungsanalyse evaluieren, inwiefern graduelle oder abrupte Änderungen aufgetreten sind.

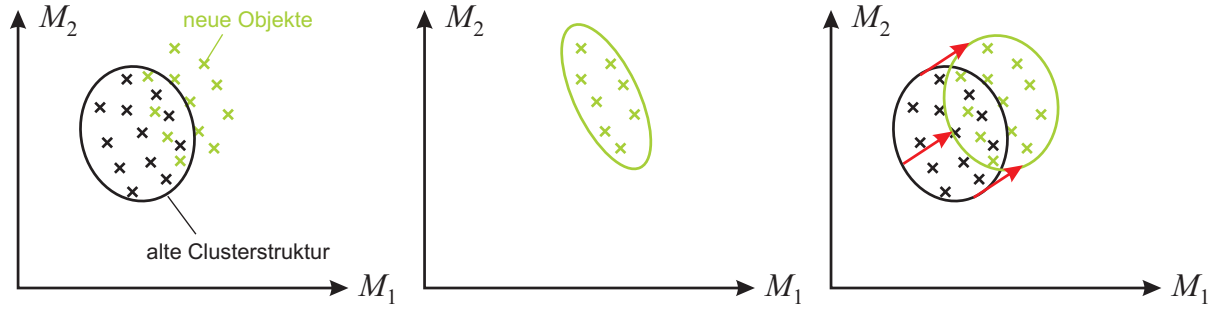
Bevor eine Untersuchung potentieller Veränderungen durchgeführt werden kann, wird zunächst analysiert, inwieweit Objekte durch Cluster absorbiert werden. Als Grenzwert der Absorbierung wird ein Parameter α^{A} gewählt, anhand dessen eine Clusterzuordnung erfolgt (vgl. Definition 4.1). Als mögliche Werte für α^{A} haben sich kleine Werte als sinnvoll erwiesen. Dieses Vorgehen bietet unter anderem den Vorteil, dass Objekte, die sich im direkten Umfeld eines Clusters befinden und so eine mögliche Verschiebung des Clusters implizieren, durch dieses Cluster absorbiert und damit in die Untersuchung eingeschlossen werden (vgl. Abschnitt 5.3). Generell ist die Zuordnung bei kleinen α^{A}-Werten stärker, d.h., mehr Objekte werden einem Cluster zugeordnet. Dies führt dazu, dass eine genauere Analyse möglich ist, da weniger Objekte fälschlicherweise als nicht zugeordnet identifiziert werden.

Durch die Absorbierung ist es möglich, Veränderungen innerhalb einzelner Cluster zu untersuchen sowie nicht absorbierte Objekte zu analysieren, so dass Aufschluss über verschiedene graduelle und abrupte Veränderungen gegeben werden kann. Wie potentielle Veränderungen daran gemessen werden können, wird in den folgenden Abschnitten ausführlich dargestellt.

5.3. Graduelle Veränderungen

Bevor abrupte Veränderungen innerhalb einer Clusterstruktur aufgedeckt werden können und eine Anpassung des zugrundeliegenden Musters vorgenommen werden kann, erfolgt zunächst eine Analyse gradueller Veränderungen bzgl. der aktuell relevanten Objekte. Die Erfassung gradueller vor der Untersuchung abrupter Veränderungen ist sinnvoll, da diese Art an Veränderungen einerseits als Indikator für spätere abrupte Veränderungen dienen kann (vgl. Tabelle 5.2), andererseits sich durch ihr Aufdecken die Untersuchung möglicher abrupter Veränderungen an verschiedenen Stellen innerhalb der Struktur jedoch erübrigen kann. Das in Abbildung 5.2 dargestellte Beispiel soll diesen Umstand verdeutlichen: Abbildung 5.2a zeigt die aus vorherigen Analysen bekannte Clusterstruktur sowie neu hinzugekommene Objekte. In Abbildung 5.2b sind die neuen, nicht zugeordneten Objekte nach Absorbierung der Objekte auf Basis der bekannten Clusterstruktur und des Grenzwertes α^{A} abgebildet. Erfolgte direkt im Anschluss eine Analyse der abrupten Veränderungen, so ergäbe diese möglicherweise zunächst ein direkt neben dem bereits bestehenden neu entstandenes Cluster. Erst eine anschließende Untersuchung gäbe Aufschluss darüber, dass sich das potentiell neue Cluster im direkten Umfeld eines bestehenden befände, somit an dieser Stelle entsprechend kein eigenständiges Cluster vorhanden wäre. Wird hingegen vorab das Auftreten gradueller Unterschiede untersucht, kann eine Verschiebung des bereits bekannten Clusters durch Fokussierung auf die aktuellen Objekte aufgedeckt werden (vgl. Abbildung 5.2c). Durch die Verschiebung wird verhindert, dass ein potentiell neues Clus-

[17] Relevante Objekte sind dabei im Zeitintervall seit dem letzten Untersuchungszeitpunkt, d.h. in den letzten Δt Zeiteinheiten neu aufgetretene Objekte oder Objekte, die sich seit der letzten Analyse verändert haben.



(a) Alte Clusterstruktur und neue Objekte

(b) Mögliches neues Cluster nach Absorbierung

(c) Verschiebung des alten Clusters: kein neues Cluster möglich

Abbildung 5.2.: Verschiebung eines Clusters

ter an dieser Stelle erkannt wird, da keine ausreichende Anzahl neu hinzugekommener Objekte vorhanden ist, die nicht bereits durch ein Cluster absorbiert werden (vgl. Abschnitt 5.4).

In der Praxis sind verschiedene graduelle Veränderungen denkbar, die für sich genommen bereits relevante Unterschiede zwischen verschiedenen Zeitpunkten anzeigen, weiterhin aber auch als Indikator zukünftiger abrupter Veränderungen dienen können. Im Folgenden wird auf die einzelnen Arten gradueller Veränderungen detailliert eingegangen und das zugehörige Vorgehen jeweils an einem Beispiel veranschaulicht.

5.3.1. Veränderung der Clusterposition

Die Veränderung der Clusterposition, d.h. bzgl. der Positionierung des Clusterprototypen sowie der Clusterausrichtung, wird auch als *Clusterdrift* bezeichnet (vgl. Angstenberger, 2000, S. 125). Im Marketingkontext kommt dieser Art der Veränderung besondere Bedeutung zu. Durch das Aufdecken des Wanderns eines Clusters und damit eines Marktsegments lässt sich unter anderem verdeutlichen, auf welche Weise sich die das entsprechende Segment charakterisierenden elementaren Eigenschaften über die Zeit hinweg verändern können, so z.B. durch Abweichungen in der Altersstruktur, steigende Einkommensverhältnisse, einen Wandel des allgemeinen Kaufverhaltens o.ä. seitens der Kunden. In diesem Abschnitt liegt der Fokus auf der Veränderung der Position des Clusterzentrums; die Clusterausrichtung wird im Rahmen von Abschnitt 5.3.2 zur Veränderung der clustereigenen Struktur näher betrachtet.

Um eine möglicherweise erfolgte Verschiebung eines Clusters zu eruieren, wird die Position des Clusterzentrums anhand der relevanten Objekte aus dem aktuell hinzukommenden Zeitintervall untersucht. Zur Anpassung des Clusterzentrums führen Crespo und Weber (2005) eine Updateregel ein, die auf einer Kombination der bekannten Zentren mit den Zentren der neu durch ein Cluster absorbierten Objekte basiert; die Zentren der neu hinzugekommenen Objekte werden dabei gemäß der Regel zur Berechnung eines Clusterzentrums in der klassischen Fuzzy-Clusteranalyse aus (3.4) bestimmt. Dieses Vorgehen ist jedoch nur bedingt geeignet, da auf diese Weise impliziert wird, dass die zur Berechnung der neuen Zentren nach (3.4) benötigten Zugehörigkeitsgrade der neuen Objekte, die auf Basis der alten Clusterstruktur geschätzt wurden, auch zum aktuellen Untersuchungszeitpunkt gültig sind. Im Falle einer probabilistischen Analyse er-

scheint dieser Ansatz aufgrund der wahrscheinlichkeitsbasierten Zugehörigkeitsgrade sinnvoll. Im possibilistischen Fall bekommen neue Objekte bei einem Clusterdrift jedoch entsprechend geringe Zugehörigkeitsgrade und beeinflussen dadurch die Schätzung des Zentrums der neuen Daten. Abbildung 5.2 verdeutlicht diesen Umstand: Würde die Verschiebung des Prototyps auf Basis der Zugehörigkeitsgrade zum alten Cluster bestimmt (vgl. Abbildung 5.2a), erhielten die neuen Objekte aufgrund ihres Abstands zum alten Clusterzentrum ein vergleichsweise geringes Gewicht; die Verschiebung (Abbildung 5.2c) würde nicht deutlich. Daher wird im Folgenden von einer Gewichtung der neuen Objekte über ihre Zugehörigkeitsgrade abgesehen; dieser Vorgang führt zu einer Anpassung der Updateregeln gegenüber Crespo und Weber (2005). Auf diese Weise wird verhindert, dass neue Objekte, die die aktuelle Entwicklung widerspiegeln, unter Umständen ein vergleichsweise geringes Gewicht erhalten, da die Zugehörigkeitsgrade auf Basis des nicht länger aktuellen Clusterprototyps bestimmt werden.

Das Zentrum der neu hinzugekommenen Objekte, die durch ein Cluster i mit Prototyp C_i absorbiert werden, wird somit folgendermaßen bestimmt:

$$\vec{v}_i^{neu} = \frac{\sum_{\vec{x}_j \in \left[U_i^{neu}\right]_{\alpha^{\mathrm{A}}}} \vec{x}_j}{n_i^{\alpha^{\mathrm{A}} neu}}, \tag{5.1}$$

wobei der Schnitt $\left[U_i^{neu}\right]_{\alpha^{\mathrm{A}}}$ die im aktuellen Zeitintervall neu hinzugekommenen, durch Cluster i absorbierten Objekte enthält; $n_i^{\alpha^{\mathrm{A}} neu}$ gibt die Anzahl dieser Objekte an.

Die Schätzung des aktuellen Clusterzentrums $\hat{\vec{v}}_i$ erfolgt gemäß Crespo und Weber (2005) anhand einer gewichteten Kombination des mit Hilfe von (5.1) bestimmten neuen Zentrums mit dem im vorherigen Zeitintervall bestimmten Clusterzentrum $\vec{v}_i$:

$$\hat{\vec{v}}_i = \left(1 - \gamma_i^{\vec{v}}\right) \vec{v}_i + \gamma_i^{\vec{v}} \vec{v}_i^{neu}, \tag{5.2}$$

wobei $\gamma_i^{\vec{v}}$ das Gewicht für den Einfluss neuer Objekte auf die Verschiebung des alten Clusterzentrums angibt. Nach Crespo und Weber (2005) soll $\gamma_i^{\vec{v}}$ dabei den Anteil der neuen, durch das Cluster absorbierten Objekte an allen im aktuellen Zeitfenster absorbierten Objekten widerspiegeln, d.h.

$$\gamma_i^{\vec{v}} = \frac{\sum_{\vec{x}_j \in \left[U_i^{neu}\right]_{\alpha^{\mathrm{A}}}} u_{ij}}{\sum_{\vec{x}_k \in [U_i]_{\alpha^{\mathrm{A}}}} u_{ik}}. \tag{5.3}$$

Beim Update der Clusterposition erfolgt gemäß (5.2) kein Ausschluss veralteter Objekte; diese erhalten indirekt Einfluss über die Verwendung der alten $\vec{v}_i$. Zur Bestimmung der $\gamma_i^{\vec{v}}$ werden hingegen nur die im aktuellen Zeitintervall in $[U_i]_{\alpha^{\mathrm{A}}}$ enthaltenen Daten berücksichtigt, veraltete Daten entfallen. Auf diese Weise soll eine adäquate Gewichtung der einzelnen Zentren bei ihrer Kombination gemäß (5.2) stattfinden. Als Folge ergäbe sich für den Spezialfall $\tau = \Delta t$, d.h. bei der Analyse einzelner Zeitfenster, dass als Schätzung der neuen Zentren die ungewichteten Zentren der neu hinzugekommen Objekte $\vec{v}_i^{neu}$ verwendet würden, da in diesem Fall $\left[U_i^{neu}\right]_{\alpha^{\mathrm{A}}} = [U_i]_{\alpha^{\mathrm{A}}}$ und damit $\gamma_i^{\vec{v}} = 1$ gälte.

Im Gegensatz zur Schätzung der $\vec{v}_i^{neu}$ gemäß (5.1) erfolgt die Berechnung der clusterspezifischen Gewichte unter Berücksichtigung der anhand der alten Clusterstruktur bestimmten

Zugehörigkeitsgrade. Bei der Wanderung eines Clusterzentrums kann eine generelle Auswirkung auf die Zugehörigkeitsgrade unterstellt werden, da die absorbierten Objekte diesem allgemeinen Drift unterliegen. Ist jedoch keine allgemeine Entwicklung vorhanden, sondern sind lediglich leichte Schwankungen Ursache einer möglichen Veränderung, so sollte das neue Zentrum $\vec{v}_i^{neu}$ bei der Schätzung der $\hat{\vec{v}}_i$ ein vergleichsweise geringes Gewicht erhalten. Das Einhalten dieser Bedingung wird bei der Berücksichtigung der Zugehörigkeitsgrade gewährleistet. Dennoch ist das Update des Clusterzentrums auf Basis der clusterspezifischen Gewichte $\gamma_i^{\vec{v}}$ gemäß (5.3) bei einem allgemeinen Clusterdrift nicht immer ausreichend, da sich der indirekte Einfluss veralteter Daten durch die Verwendung der alten $\vec{v}_i$ als zu stark erweist; dieser kann durch eine Erhöhung der Gewichtung neu hinzugekommener Objekte bei der Bestimmung der $\gamma_i^{\vec{v}}$ verringert werden und damit zu folgender Anpassung der Bestimmung clusterspezifischer Gewichtungsparameter führen:

$$\begin{aligned} \gamma_i^{\vec{v}} &= \frac{\gamma^{\vec{v}} \sum_{\vec{x}_j \in \left[U_i^{neu}\right]_{\alpha^{\mathrm{A}}}} u_{ij}}{\gamma^{\vec{v}} \sum_{\vec{x}_k \in \left[U_i^{neu}\right]_{\alpha^{\mathrm{A}}}} u_{ik} + \sum_{\vec{x}_l \in [U_i]_{\alpha^{\mathrm{A}}} \setminus \left[U_i^{neu}\right]_{\alpha^{\mathrm{A}}}} u_{il}} \\ &= \frac{\gamma^{\vec{v}} \sum_{\vec{x}_j \in \left[U_i^{neu}\right]_{\alpha^{\mathrm{A}}}} u_{ij}}{\left(\gamma^{\vec{v}} - 1\right) \sum_{\vec{x}_k \in \left[U_i^{neu}\right]_{\alpha^{\mathrm{A}}}} u_{ik} + \sum_{\vec{x}_l \in [U_i]_{\alpha^{\mathrm{A}}}} u_{il}}, \end{aligned} \tag{5.4}$$

wobei $\gamma^{\vec{v}}$ ein allgemeiner Gewichtungsparameter für den Einfluss der neu hinzugekommen Objekte ist. Auf diese Weise kann die Gewichtung der neuen Objekte je nach Zeitmodell angepasst werden; bei Gleichgewichtung mit $\gamma^{\vec{v}} = 1$ ergibt sich eine Gewichtung gemäß (5.3) nach Crespo und Weber (2005). Eine geeignete Wahl von $\gamma^{\vec{v}}$ ist von verschiedenen Größen abhängig:

- τ: Je höher τ, desto höher sollte die gewählte Gewichtung sein, um den Fokus auf neuere Daten zu legen. Je größer τ, d.h. je länger das betrachtete Zeitintervall, desto älter und damit weniger aktuell sind die Daten, die indirekt über die alten Clusterzentren in die Berechnung eingehen. Durch eine entsprechend höhere Gewichtung der neuen Daten können allgemeine Trends dennoch aufgedeckt werden.
- $\frac{\tau}{\Delta t}$: Je höher das Verhältnis zwischen Zeitfensterlänge und Analysefrequenz, desto höher sollte die Gewichtung sein, sofern nur geringe Veränderungen erwartet werden. In der Praxis ist davon auszugehen, dass keine abrupten Sprünge stattfinden. Hierbei ist jedoch zu beachten, dass die Periodenlänge nicht zu groß gesetzt sein darf.
- Δt: Je höher die Analysefrequenz Δt, desto höher sollte die Gewichtung sein, falls starke Veränderungen erwartet werden.
- α^{A}: Je kleiner der Grenzwert der Absorbierung, desto geringer sollte dich Gewichtung sein, da die Schätzung des neuen Zentrums ungewichtet erfolgt.

Generell ist die Wahl von $\gamma^{\vec{v}}$ kontextabhängig, da bei starken Verschiebungen einzelne Objekte den Absorptionsbereich verlassen oder nur sehr kleine Zugehörigkeitsgrade bekommen. Empirisch hat sich bei Zeitfenstern der Länge $\tau \leq 4$ unabhängig von den übrigen Parametern $\gamma^{\vec{v}} = \tau$ als geeignet erwiesen. Bei größeren Zeitfenstern sind die übrigen Parameter in die Wahl des Gewichtungsparameters einzubeziehen.

Anstelle der Zugehörigkeitsgrade, die auf Basis des veralteten Clusterprototyps geschätzt werden, ist bei der Bestimmung des Gewichts $\gamma_i^{\vec{v}}$ die Verwendung der absoluten Anzahl der absorbierten Objekte zum jeweiligen Zeitpunkt möglich. Dieses Vorgehen bietet den Vorteil, dass keine veralteten Zugehörigkeitsgrade einbezogen werden, die gerade bei starken Schwankungen ins Gewicht fielen. Allerdings wird die Schätzung allgemein deutlich anfälliger gegenüber

einfachen zufälligen Schwankungen innerhalb eines Datensatzes; deswegen sollte im Falle der Schätzung der Clusterposition die Verwendung der Zugehörigkeitsgrade vorgezogen werden.

Wird eine Verschiebung des Clusterzentrums aufgedeckt, muss evaluiert werden, inwiefern diese Änderung signifikant ist und nicht nur das Produkt zufälliger Schwankungen darstellt. Hierzu werden die Fuzzy-Varianzen der einzelnen Dimensionen auf Basis der alten Clusterstruktur bestimmt (vgl. (3.8)) und dimensionsweise mit den Veränderungen der einzelnen Clusterzentren verglichen. Eine Veränderung wird adaptiert, falls gilt

$$\Delta v_{il} \geq \beta^{\vec{v}} \sigma^2_{ill}, \tag{5.5}$$

wobei

- Δv_{il}: Veränderung des Clusterzentrums von Cluster i bzgl. Dimension l im Vergleich zur vorherigen Periode,
- σ^2_{ill}: Fuzzy-Varianz von $[U_i]_{\alpha^{\mathrm{A}}}$ der l-ten Dimension in alter Clusterstruktur,
- $\beta^{\vec{v}} \in [0, 1]$: Parameter zur Bestimmung des geforderten Anteils.

Die Wahl von $\beta^{\vec{v}}$ richtet sich nach der benötigten Sensibilität gegenüber Veränderungen und erfolgt kontextabhängig. Die isolierte Betrachtung der Veränderung bzgl. einzelner Dimensionen unabhängig von der kombinierten Veränderung ist ausreichend, da insbesondere im Marketingkontext eine Veränderung der das Cluster charakterisierenden Eigenschaften aufzudecken ist. Erst bei der anschließenden Interpretation der Veränderung ist die Kombination einzelner Änderungen relevant. Im Anschluss an das Reclustering der Objekte auf Basis aller ermittelten Unterschiede, d.h. sowohl der graduellen als auch der abrupten, muss eine Identifikation der Clustertrajektorien aller Cluster erfolgen, um den Bewegungspfad des Clusters nachvollziehen zu können. Die Identifikation sollte sinnvollerweise erst nach der erneuten Clusteranalyse erfolgen, da die Frage, welche Position sich zur Präsentation der aktuellen Struktur eignet, zunächst geklärt werden muss. Mit Hilfe der Clustertrajektorien können Trends veranschaulicht und zukünftige Entwicklungen auch bzgl. Clustervereinigung o.ä. vorhergesagt werden; das genaue Vorgehen wird in den einzelnen Abschnitten zu abrupten Veränderungen vorgestellt.

Das folgende Beispiel soll das Vorgehen zur Clusterverschiebung verdeutlichen. Zur Veranschaulichung wurde ein Datensatz mit drei runden Clustern mit einer Varianz von jeweils 1 innerhalb der Dimensionen generiert, für die in jeder Periode 50 Objekte hinzukommen. Betrachtet werden insgesamt vier Perioden. Bei zwei der Cluster (Cluster 2 und Cluster 3 in Abbildung 5.3) ist das zur Generierung verwendete Clusterzentrum konstant, bei einem jedoch (Cluster 1) wurde das Clusterzentrum in jeder Periode um denselben Vektor $\begin{pmatrix} 0.8 \\ 0.8 \end{pmatrix}$ verschoben. Die konkreten Werte sind in Tabelle 5.3 aufgeführt.

Als Länge des Zeitfensters in der Analyse wurde $\tau = 3$ gewählt, so dass für jedes Cluster bei der Analyse 150 Objekte vorhanden sind, d.h., insgesamt werden je Zeitintervall 450 Objekte untersucht. Die Verschiebung des Zeitfensters wurde auf $\Delta t = 1$ gesetzt. Dementsprechend ist eine Analyse zum Zeitpunkt $t = 3$ sowie zum Zeitpunkt $t = 4$ möglich. Als genereller Gewichtungsparameter für den Einfluss der neuen Objekte wurde wie zuvor angegeben $\gamma^{\vec{v}} = \tau$ verwendet. Der Grenzwert der Absorbierung wurde auf $\alpha^{\mathrm{A}} = 0.125$ festgelegt. Ferner wird von einem Wert $\beta^{\vec{v}} = 0.5$ ausgegangen.

Cluster	**Zentrum** $t = 1$	**Zentrum** $t = 2$	**Zentrum** $t = 3$	**Zentrum** $t = 4$
Cluster 1	$\begin{pmatrix} 2 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 2.8 \\ 4.8 \end{pmatrix}$	$\begin{pmatrix} 3.6 \\ 5.6 \end{pmatrix}$	$\begin{pmatrix} 4.4 \\ 6.4 \end{pmatrix}$
Cluster 2	$\begin{pmatrix} 11 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 11 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 11 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 11 \\ 12 \end{pmatrix}$
Cluster 3	$\begin{pmatrix} 12 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 3.5 \end{pmatrix}$

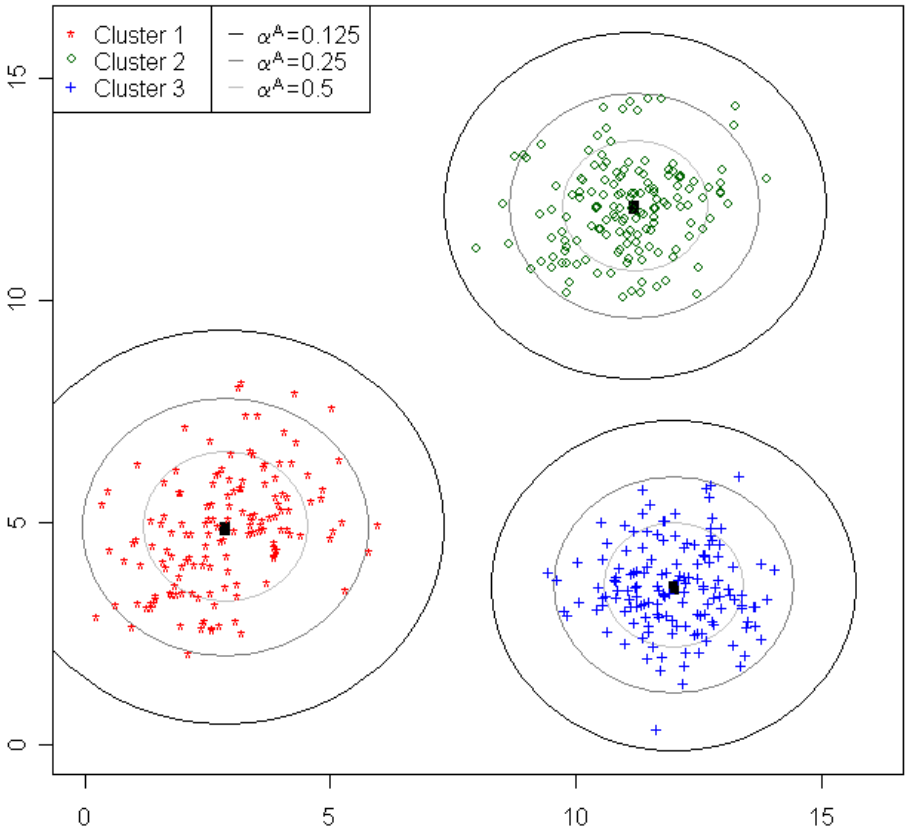
Tabelle 5.3.: Vorgegebene Clusterzentren je Periode

In Abbildung 5.3 sind die Ergebnisse des Fuzzy-C-Means auf Basis der Dreiecksbeziehung der Distanzen aufgezeigt (vgl. Abschnitt 4.3.3, Algorithmus 4.7). Abbildung 5.3a zeigt die Clusterstruktur der Daten des ersten Zeitfensters, d.h. der Analyse zum Zeitpunkt $t = 3$. Die konkreten Werte sind in Tabelle 5.4 angegeben. Die größere Ausdehnung des ersten Clusters (η_1) im Vergleich zu den übrigen Clustern und die damit einhergehenden höheren Varianzen bzgl. der einzelnen Dimensionen (σ^2_{111} bzw. σ^2_{122}) ergeben sich aus der ovalen Clusterform, die durch die Verschiebung des Clusterzentrums zur Generierung der Zufallswerte begründet ist.

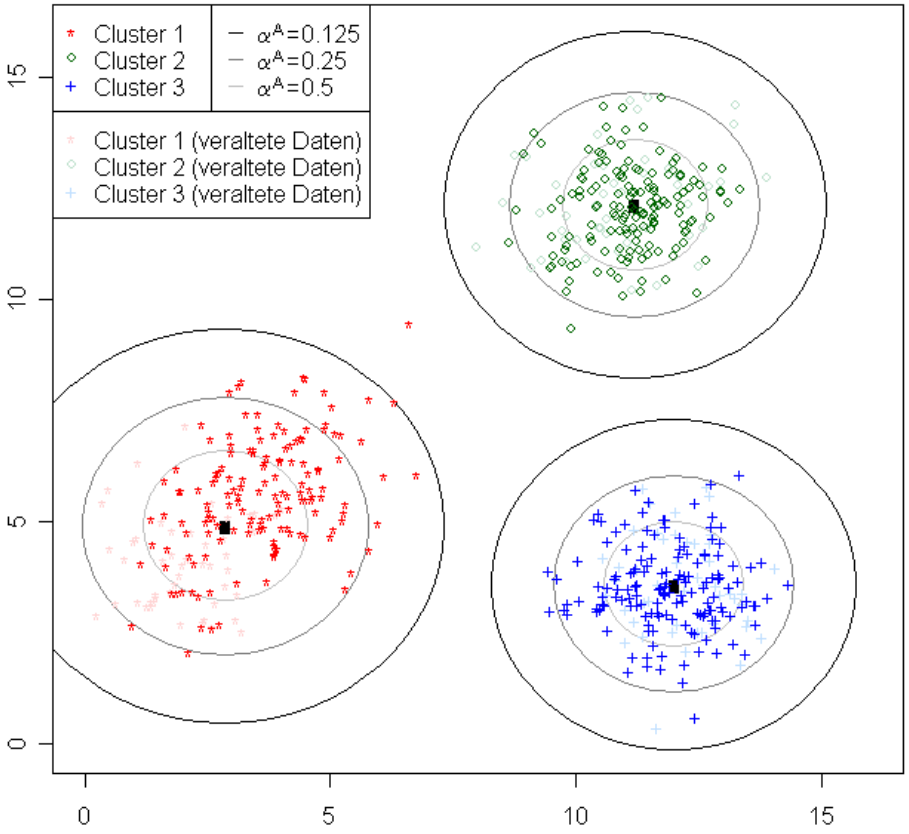
Cluster	**Zentrum** $\vec{v}_i$	**Ausdehnung** (η_i)	**Fuzzy-Varianzen** (σ^2_{ill})
Cluster 1	$\vec{v}_1 = \begin{pmatrix} 2.879 \\ 4.910 \end{pmatrix}$	$\eta_1 = 2.808$	$\sigma^2_{111} = 0.469$ $\sigma^2_{122} = 0.477$
Cluster 2	$\vec{v}_2 = \begin{pmatrix} 11.203 \\ 12.133 \end{pmatrix}$	$\eta_2 = 2.157$	$\sigma^2_{211} = 0.343$ $\sigma^2_{222} = 0.320$
Cluster 3	$\vec{v}_3 = \begin{pmatrix} 11.996 \\ 3.596 \end{pmatrix}$	$\eta_3 = 1.973$	$\sigma^2_{311} = 0.345$ $\sigma^2_{322} = 0.309$

Tabelle 5.4.: Ergebnisse der Clusteranalyse zum Zeitpunkt $t = 3$

In Abbildung 5.3 sind neben den Clusterzentren die Clustergrenzen für verschiedene Grenzwerte der Absorbierung $\alpha^{\mathrm{A}} \in \{0.125, 0.25, 0.5\}$ gegeben. Bei der isolierten Betrachtung eines Zeitintervalls der Länge τ erscheint bei einer eindeutigen Separierung der Cluster ein Grenzwert von $\alpha^{\mathrm{A}} = 0.25$ als ausreichend, um die Clusterstruktur darzustellen. Die Wahl eines geringeren Wertes erweist sich jedoch als notwendig, um Veränderungen in der Struktur aufdecken zu können. Dies wird in Abbildung 5.3b deutlich: Durch die Verschiebung des Zentrums werden viele der neu hinzugekommenen Objekte bei einem Grenzwert von $\alpha^{\mathrm{A}} = 0.25$ nicht absorbiert, so dass sie bei einer Analyse der graduellen Veränderungen nicht berücksichtigt werden würden. Bei der Verwendung von $\alpha^{\mathrm{A}} = 0.125$ werden sie jedoch in die Analyse einbezogen und verdeutlichen damit die Veränderung. Durch die Verwendung von (5.4) bei der Schätzung der neuen Zentren zur Gewichtung des Einflusses des neuen Clusterzentrums ist der Einfluss der neuen Daten hoch genug, um die Veränderung des Clusterzentrums des ersten Clusters hervorzuheben. Die vollständigen Ergebnisse der Analyse zur Veränderung der Clusterposition sind in Tabelle 5.5 gegeben; grafisch ist die Anpassung in Abbildung 5.3c dargestellt. Es wird deutlich, dass durch die Analyse die Positionsänderung des ersten Clusters sichtbar wird. Bei den übrigen Clustern sind die neuen Clusterzentren kaum von den alten zu unterscheiden. Diese

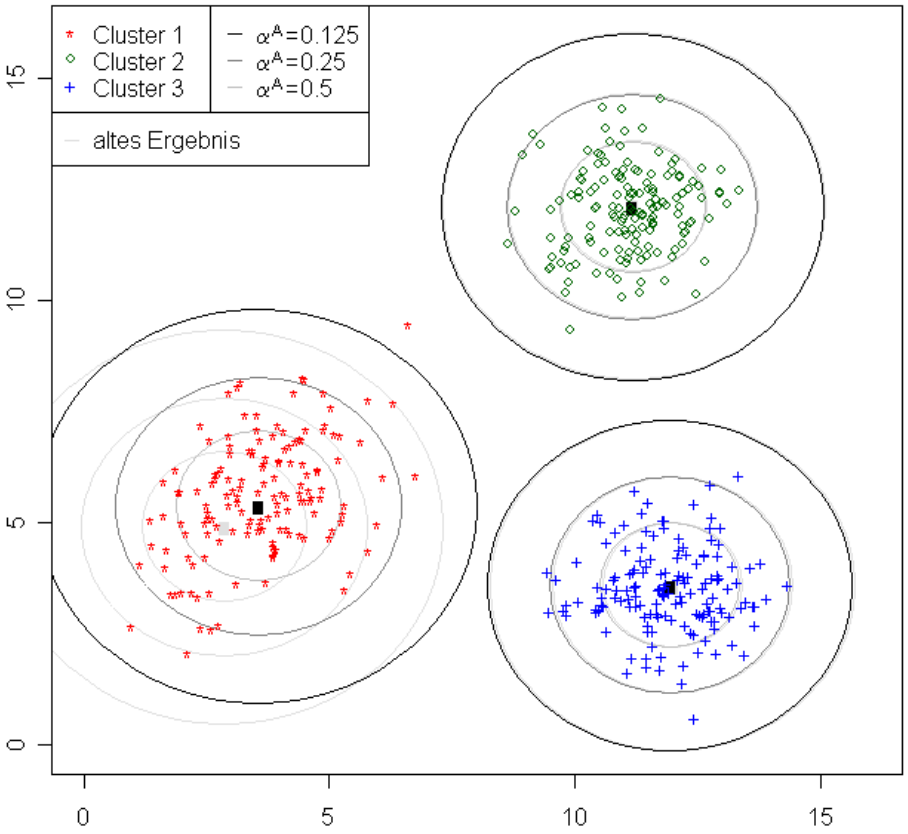


(a) Clusterstruktur für den Zeitpunkt $t = 3$ (Perioden 1-3)

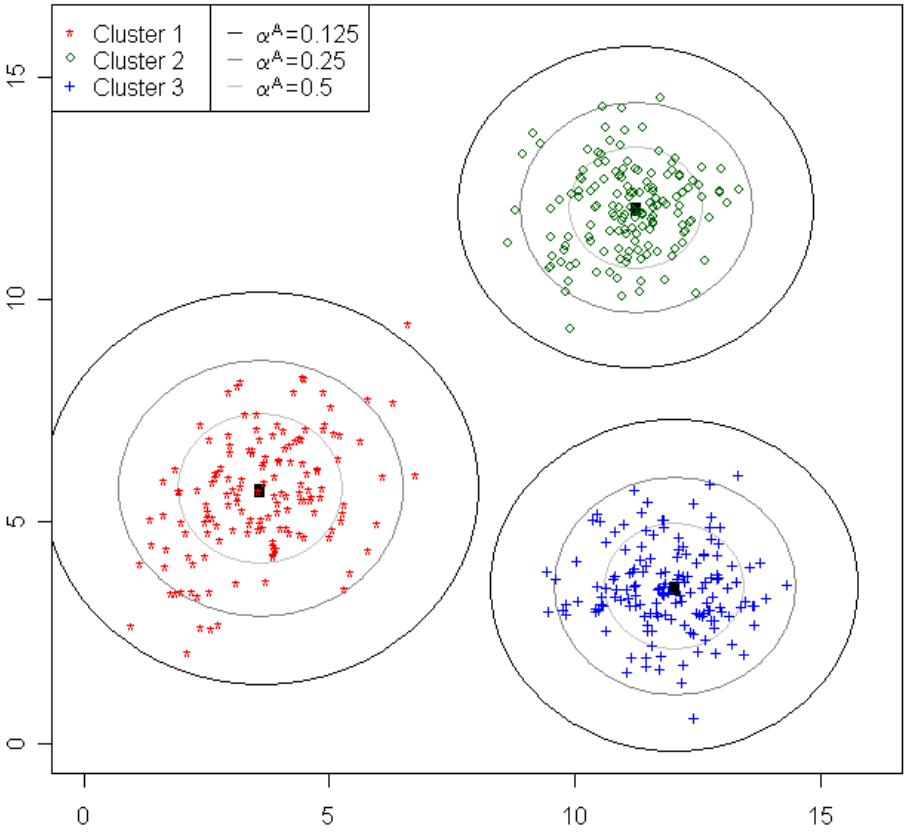


(b) Clusterstruktur $t = 3$ für neue Daten aus $t = 4$ (Perioden 2-4)

Abbildung 5.3.: Beispiel zur Verschiebung eines Clusters



(c) Verschiebung der Clusterzentren



(d) Clusterstruktur für den Zeitpunkt $t = 4$ (Perioden 2-4)

Abbildung 5.3.: Beispiel zur Verschiebung eines Clusters

Cluster	**Gewicht** $\gamma_i^{\vec{v}}$	**Zentrum der neuen Daten** $(\vec{v}_i^{neu})$	**Schätzung neues Zentrum** $(\hat{\vec{v}}_i)$	**Änderung gegenüber altem Zentrum** $(\Delta\vec{v}_i)$
Cluster 1	$\gamma_1^{\vec{v}} = 0.491$	$\vec{v}_1^{neu} = \begin{pmatrix} 4.276 \\ 6.391 \end{pmatrix}$	$\hat{\vec{v}}_1 = \begin{pmatrix} 3.565 \\ 5.391 \end{pmatrix}$	$\Delta\vec{v}_1 = \begin{pmatrix} 0.686 \\ 0.480 \end{pmatrix}$
Cluster 2	$\gamma_2^{\vec{v}} = 0.595$	$\vec{v}_2^{neu} = \begin{pmatrix} 11.154 \\ 12.041 \end{pmatrix}$	$\hat{\vec{v}}_2 = \begin{pmatrix} 11.174 \\ 12.102 \end{pmatrix}$	$\Delta\vec{v}_2 = \begin{pmatrix} -0.029 \\ -0.031 \end{pmatrix}$
Cluster 3	$\gamma_3^{\vec{v}} = 0.587$	$\vec{v}_3^{neu} = \begin{pmatrix} 11.930 \\ 3.534 \end{pmatrix}$	$\hat{\vec{v}}_3 = \begin{pmatrix} 11.958 \\ 3.575 \end{pmatrix}$	$\Delta\vec{v}_3 = \begin{pmatrix} -0.039 \\ -0.021 \end{pmatrix}$

Tabelle 5.5.: Ergebnisse der Analyse zur Veränderung der Clusterposition

Veränderungen werden gemäß (5.5) entsprechend verworfen. Dabei scheint die Wahl von $\beta^{\vec{v}}$ aufgrund der sehr guten Separierung der einzelnen Cluster im gegebenen Beispiel nahezu beliebig, da die Veränderungen beim wandernden Cluster in jedem Fall höher als die Fuzzy-Varianzen der einzelnen Dimensionen (vgl. Tabelle 5.4), die Schwankungen der übrigen Cluster dagegen vernachlässigbar gering sind.

Zum Vergleich wird abschließend das Ergebnis der Analyse zum Zeitpunkt $t = 4$ bei zufälliger Initialisierung des Algorithmus angegeben. Abbildung 5.3d stellt das Ergebnis grafisch dar; die Werte sind zusammen mit der Differenz zu den geschätzten Werten in Tabelle 5.6 aufgeführt. Durch die Analyse der Positionsänderung der Clusterzentren wird eine ausreichende Schätzung der Clusterpositionen möglich.

Cluster	**Zentrum** $\vec{v}_i$	**Ausdehnung** (η_i)	**Fuzzy-Varianzen** (σ_{ill}^2)	**Differenz der Zentren** $(\vec{v}_i - \hat{\vec{v}}_i)$
Cluster 1	$\vec{v}_1 = \begin{pmatrix} 3.610 \\ 5.749 \end{pmatrix}$	$\eta_1 = 2.795$	$\sigma_{111}^2 = 0.443$ $\sigma_{122}^2 = 0.472$	$\vec{v}_1 - \hat{\vec{v}}_1 = \begin{pmatrix} 0.045 \\ 0.358 \end{pmatrix}$
Cluster 2	$\vec{v}_2 = \begin{pmatrix} 11.243 \\ 12.079 \end{pmatrix}$	$\eta_2 = 1.862$	$\sigma_{211}^2 = 0.271$ $\sigma_{222}^2 = 0.288$	$\vec{v}_2 - \hat{\vec{v}}_2 = \begin{pmatrix} 0.069 \\ -0.023 \end{pmatrix}$
Cluster 3	$\vec{v}_3 = \begin{pmatrix} 12.032 \\ 3.558 \end{pmatrix}$	$\eta_3 = 1.997$	$\sigma_{311}^2 = 0.353$ $\sigma_{322}^2 = 0.291$	$\vec{v}_3 - \hat{\vec{v}}_3 = \begin{pmatrix} 0.075 \\ -0.017 \end{pmatrix}$

Tabelle 5.6.: Ergebnisse der Clusteranalyse zum Zeitpunkt $t = 4$

5.3.2. Veränderung der clustereigenen Struktur

Neben dem Clusterzentrum, das die generelle Position eines Clusters angibt, besitzt ein Cluster weitere Eigenschaften, durch die seine clustereigene Struktur festgelegt ist. Im Falle kugelförmiger Cluster sind dies insbesondere die Clusterdichte und das Volumen, bei ellipsoiden Clustern ist außerdem die generelle Clusterausrichtung relevant, die die unterschiedliche Ausdehnung innerhalb einzelner Dimensionen berücksichtigt. Im Rahmen der Untersuchung gradueller Unterschiede müssen auch Veränderungen bzgl. der clustereigenen Struktur Beachtung finden. Folgende Veränderungen können anhand von Clusterdichte und -volumen aufgedeckt werden:

1. Veränderung der Objektzahl: Die Anzahl der durch ein Cluster absorbierten Objekte gibt Aufschluss darüber, ob ein Cluster wächst bzw. sich verdichtet oder ob die Entwicklung rückläufig ist. Nimmt beispielsweise die Objektzahl und damit einhergehend die Dichte innerhalb eines Clusters stetig ab, so kann dies ein Anzeichen für eine bevorstehende Clusterelimination sein (vgl. Tabelle 5.2).

2. Veränderung der Clusterausdehnung: Die Ausdehnung eines Clusters hängt mit der Varianz dieses Clusters innerhalb der einzelnen Dimensionen und den zugehörigen Kovarianzen zusammen. Neben einer allgemeinen Veränderung des Clustervolumens, bei der alle von einem Cluster absorbierten Objekte demselben Trend bzgl. der Veränderung der Clusterausdehnung unterliegen, ist bei ellipsoiden Clustern auch die Veränderung der Ausdehnung innerhalb einzelner Dimensionen möglich, die eine Veränderung der Clusterausrichtung und damit einhergehend eine Rotation des Clusters verursacht. Dies ist beispielsweise dann von Bedeutung, wenn herausgestellt werden soll, dass bestimmte Eigenschaften für ein Segment an Bedeutung gewinnen, während die Relevanz anderer Eigenschaften abnimmt.

Ein Spezialfall der Analyse gradueller Unterschiede innerhalb der clustereigenen Struktur tritt dann auf, wenn sich die Verteilung innerhalb eines Clusters dahingehend verändert, dass an verschiedenen Stellen Regionen höherer Dichte entstehen, während das eigentliche Clusterzentrum im Extremfall verwaist. Da es in diesem Fall zum gleichzeitigen Auftreten verschiedener gradueller Veränderungen kommt, wird er jedoch gesondert betrachtet werden. Die Verwaisung des Clusterzentrums kann eine bevorstehende Clusterteilung implizieren.

Die Untersuchung der Veränderung der clustereigenen Struktur erfolgt im Anschluss an die Clusterverschiebung, da auf diese Weise ausgeschlossen werden kann, dass die Variationen in Dichte und Volumen durch eine Verschiebung des Clusters hervorgerufen wurden. Dieser Umstand soll an dem in Abbildung 5.2 angeführten Beispiel (S. 78) zur Clusterverschiebung verdeutlicht werden. Analysierte man das Cluster nur auf Basis der neuen Objekte in Abbildung 5.2a, ergäbe die Untersuchung einen Rückgang bzgl. Dichte und Volumen, da sich lediglich am Rand des Clusters neue Objekte befänden. Erfolgt die Evaluierung der Dichte jedoch nach der Clusterverschiebung, wird deutlich, dass diesbezüglich keine relevante Veränderung vorhanden ist (vgl. Abbildung 5.2c).

Zur Evaluation von Veränderungen der allgemeinen Clusterstruktur können unterschiedliche Maße zur Bewertung eines Clusters hinzugezogen werden. Die absolute Kardinalität $n_i^{\alpha^{\mathrm{A}}}$ eines Clusters bzw. des α-Schnitts des Clusters gibt bereits Aufschluss über eine mögliche Veränderung der Anzahl der Objekte, die durch ein Cluster absorbiert werden. Weitere Anhaltspunkte, insbesondere bzgl. der Dichte innerhalb eines Clusters, liefert die Fuzzy-Kardinalität, die die Zugehörigkeitsgrade innerhalb des Clusters in die Bestimmung einbezieht (vgl. (3.16)). Die Betrachtung der Fuzzy-Kardinalität basiert auf der Untersuchung der Veränderung der Partitionsdichte als lokales Gütemaß als Teil der in Abschnitt 3.5 vorgestellten mittleren Partitionsdichte aus (3.16). Da das Clustervolumen der vorangegangenen Periode angenommen wird, d.h., da die Kovarianzmatrizen unverändert sind, kann dabei auf die Normierung über das Clustervolumen verzichtet werden; dadurch reduziert sich die Bestimmung der Partitionsdichte eines Clusters auf die Fuzzy-Kardinalität von $Y_i = \left\{\vec{x}_j \middle| \left(\vec{x}_j - \vec{v}_i\right)^{\mathrm{T}} \Sigma_i^{-1} \left(\vec{x}_j - \vec{v}_i\right) < 1, j \in \{1, \ldots, n\}\right\}$. Aufgrund der Eigenschaft, dass Σ_i symmetrisch und positiv semidefinit ist, lässt sich die Be-

stimmung der Partitionsdichte wegen der Schätzung $\eta_i = \det(\Sigma_i)^{\frac{1}{p}}$ (vgl. Krishnapuram und Keller, 1993) und der Substitution der Kovarianzmatrizen durch die positiv definite Normmatrix $A_i = \det(\Sigma_i)^{\frac{1}{p}} \Sigma_i^{-1}$ (vgl. Abschnitt 4.2.2 sowie Timm (2002), S. 20) näherungsweise reduzieren auf die Fuzzy-Kardinalität des α-Schnitts bei $\alpha^{\mathrm{A}} = 0.5$[18], indem Y_i mit

$$Y_i = \left\{ \vec{x}_j \middle| (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} \Sigma_i^{-1} (\vec{x}_j - \vec{v}_i) < 1, j \in \{1, \ldots, n\} \right\}$$

substituiert wird durch

$$\hat{Y}_i = \left\{ \vec{x}_j \middle| (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i) < \eta_i, j \in \{1, \ldots, n\} \right\}.$$

Da die η_i die Distanz vom Clusterzentrum repräsentieren, bei der Objekte einen Zugehörigkeitsgrad von 0.5 zum Cluster bekommen (vgl. Abschnitt 3.4), entspricht die Fuzzy-Kardinalität von $\hat{Y}_i$ der Fuzzy-Kardinalität des α-Schnitts bei $\alpha^{\mathrm{A}} = 0.5$. Die Verwendung der Fuzzy-Kardinalität anstelle der lokalen Partitionsdichte bietet den Vorteil, dass sie für verschiedene Werte für α^{A} mit

$$\mathrm{card}\left([U_i]_{\alpha^{\mathrm{A}}}\right) = \sum_{\vec{x}_j \in [U_i]_{\alpha^{\mathrm{A}}}} u_{ij} \tag{5.6}$$

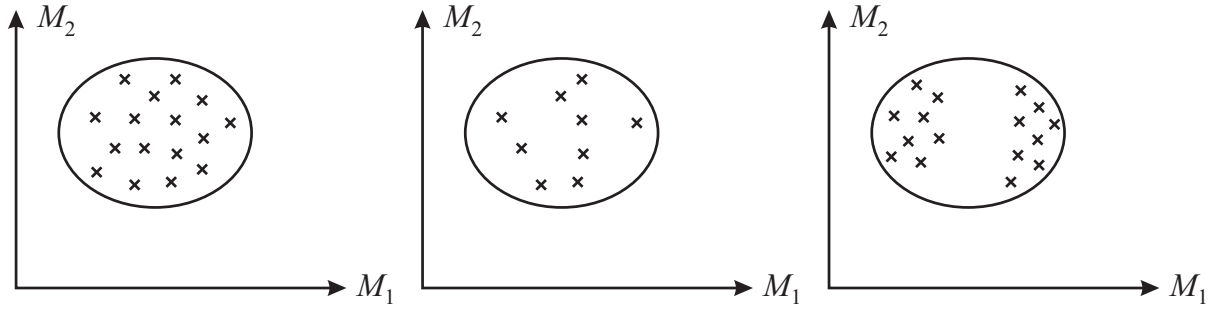
bestimmt werden kann.

Die Veränderung der Fuzzy-Kardinalität gibt bereits Aufschluss über eine Schwankung innerhalb des Clusters. So kann eine Erhöhung als Indiz für eine zunehmende Homogenisierung innerhalb des Clusters angesehen werden, entweder aufgrund einer generellen Zunahme der durch das Cluster absorbierten Objekte oder aufgrund einer allgemeinen Verringerung des Clustervolumens. Für einen Rückgang der Fuzzy-Kardinalität sind jedoch zwei Ursachen denkbar (vgl. Abbildung 5.4): Zum einen kann eine generelle Verringerung der Objektzahl innerhalb eines Clusters aufgrund eines Rückgang der Objektzahl im Allgemeinen bzw. einer Ausdehnung des Clustervolumens Auslöser einer veränderten Fuzzy-Kardinalität sein (Abbildung 5.4b), wodurch eine zukünftige Elimination des Clusters impliziert werden kann. Zum anderen kann das Zurückgehen der Fuzzy-Kardinalität durch eine Verwaisung des Clusterzentrums ausgelöst werden (Abbildung 5.4c), was unter anderem auf eine anstehende Clusterteilung hindeuten kann. Um die Veränderung der Fuzzy-Kardinalität besser beurteilen zu können, werden daher zusätzlich weitere Maße hinzugezogen. Zur Untersuchung der generellen Kompaktheit des Clusters führen Bensaid u. a. (1996) einen Kompaktheitsindex ein, der die Varianz innerhalb eines Clusters auf Basis der gewichteten quadratischen Abweichung vom Clusterzentrum und der Fuzzy-Kardinalität bestimmt:

$$\kappa_i^{\alpha^{\mathrm{A}}} = \frac{\sum_{\vec{x}_j \in [U_i]_{\alpha^{\mathrm{A}}}} u_{ij}^m d_{C_i}^2 (\vec{v}_i, \vec{x}_j)}{\mathrm{card}\left([U_i]_{\alpha^{\mathrm{A}}}\right)} \tag{5.7}$$

Je kleiner der Kompaktheitsindex $\kappa_i^{\alpha^{\mathrm{A}}}$, desto geringer ist die Fuzzy-Varianz des Clusters, d.h., desto kompakter ist das Cluster. Verwaist das Clusterzentrum beispielsweise aufgrund einer

[18] Es handelt sich hierbei um eine Annäherung und kein exaktes Ergebnis, da die η_i zu Beginn einer possibilistischen Analyse geschätzt werden, die A_i jedoch im Laufe der Analyse angepasst werden. Bei den Erweiterungen zur possibilistischen Analyse werden die clusterspezifischen Kovarianzmatrizen Σ_i jedoch nicht explizit bestimmt, weshalb eine solche Annäherung erforderlich ist (vgl. Kapitel 4).



(a) Ausgangslage (b) Verringerung der Objektzahl (c) Verwaisung des Clusterzentrums

Abbildung 5.4.: Ursachen für Rückgang der Fuzzy-Kardinalität

bevorstehenden Trennung nach und nach, so wird dies durch einen stark wachsenden Kompaktheitsindex aufgrund der zunehmenden Fuzzy-Varianz innerhalb dieses Clusters aufgedeckt.

Für eine weiterführende Evaluation struktureller Veränderungen können außerdem lokale Maße aus globalen Validitätskriterien abgeleitet werden[19]. Der in Abschnitt 3.5 eingeführte Partitionskoeffizient PC (vgl. (3.13)) kann als lokales Maß ($LPC\left([U_i]_{\alpha^{\mathrm{A}}}\right)$) zur Unterstützung hinzugezogen werden, um Unterschiede in den Zugehörigkeitsgraden und der Verteilung innerhalb eines Clusters aufzudecken, da er auf möglichst harte Clusterzuordnungen ausgerichtet ist.

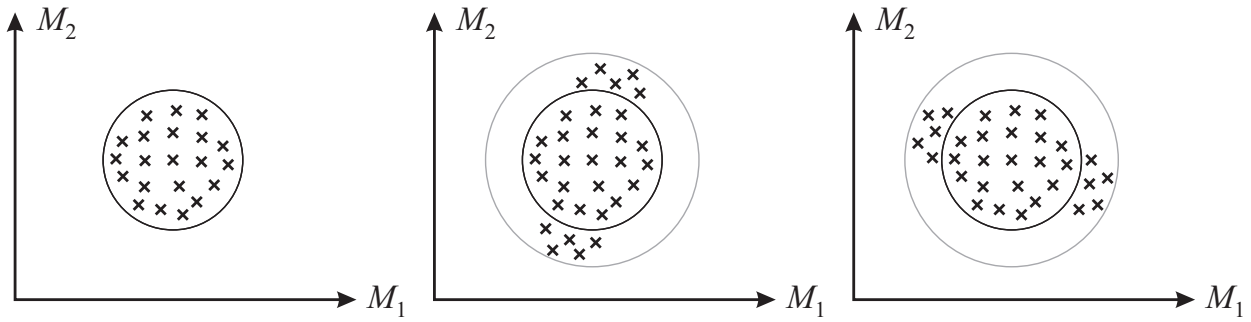
$$LPC\left([U_i]_{\alpha^{\mathrm{A}}}\right) = \frac{1}{n_i^{\alpha^{\mathrm{A}}}} \sum_{\vec{x}_j \in [U_i]_{\alpha^{\mathrm{A}}}} u_{ij}^2 \tag{5.8}$$

Bei ellipsoiden Clustern kann außerdem die Kovarianzmatrix zur Überprüfung der Veränderung innerhalb eines Clusters hinzugezogen werden. Ein Update der Fuzzy-Kovarianzmatrizen analog zu Schätzung der neuen Clusterzentren durchzuführen (vgl. Abschnitt 5.3.1), erweist sich jedoch als wenig sinnvoll, da die alte Kovarianzmatrix jeweils durch die Absorbierung auf Basis ebendieser Matrix indirekt in die Schätzung einginge. Hätte sich ein Cluster beispielsweise gedreht, könnte dies mit Hilfe der Fuzzy-Kovarianzmatrix nicht verdeutlicht werden, da ungeeignete Zugehörigkeitsgrade für die Abschätzung hinzugezogen würden. Die Anpassung der Kovarianzmatrix darf in diesem Fall jedoch nicht – wie bei der Aktualisierung der Clusterposition – ohne eine Gewichtung auf Basis der Zugehörigkeitsgrade erfolgen, da sonst die Schwankungen aufgrund der unterschiedlichen Vorgehensweisen derart groß wären, dass sie nicht länger zur Messung der Unterschiede herangezogen werden könnten. Verwendet man jedoch anstelle der Fuzzy-Kovarianzmatrix in beiden Fällen die entsprechende harte Kovarianzmatrix der absorbierten Objekte (vgl. (5.9)), wird ein Vergleich möglich.

$$\Sigma_i^{hart} = \frac{\sum_{\vec{x}_j \in [U_i]_{\alpha^{\mathrm{A}}}} \left(\vec{x}_j - \vec{v}_i\right) \left(\vec{x}_j - \vec{v}_i\right)^{\mathrm{T}}}{n_i^{\alpha^{\mathrm{A}}}} \tag{5.9}$$

[19] Bei lokaler Anwendung globaler Gütekriterien zur Evaluierung struktureller Veränderungen werden diese nicht zur generellen Bewertung eines einzelnen Clusters verwendet, sie dienen lediglich dem lokalen Vergleich und dem Aufdecken interner Veränderungen.

Dabei wird ein niedriger Grenzwert der Absorption benötigt (z.B. $\alpha^{\mathrm{A}} = 0.125$). Für größere Werte wird ein zu kleiner Ausschnitt des Clusters einbezogen, so dass tatsächliche Veränderungen wie z.B. das Drehen eines Cluster im betrachteten Bereich nicht ausreichend dargestellt werden. Abbildung 5.5 verdeutlicht diesen Umstand: Bei der Wahl eines zu großen Grenzwertes (Abbildung 5.5a) können die unterschiedlichen Clusterausrichtungen in den Abbildungen 5.5b und 5.5c nicht unterschieden werden, da der Betrachtungsraum zu klein ist.



(a) Kleiner Betrachtungsraum (b) Größerer Betrachtungsraum (1) (c) Größerer Betrachtungsraum (2)

Abbildung 5.5.: Einfluss von α^{A} auf resultierende Kovarianzmatrix

Das Update der harten Kovarianzmatrizen erfolgt analog zum Update der Clusterposition:

$$\Sigma_i^{neu} = \frac{\sum_{\vec{x}_j \in [U_i]_{\alpha^{\mathrm{A}}}} (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}}}{n_i^{\alpha^{\mathrm{A}neu}}} \tag{5.10}$$

$$\hat{\Sigma}_i^{hart} = \left(1 - \gamma_i^{\Sigma}\right) \Sigma_i^{hart} + \gamma_i^{\Sigma} \Sigma_i^{neu} \tag{5.11}$$

γ_i^{Σ} gibt das Gewicht für den Einfluss neuer Objekte auf die Anpassung der harten Kovarianzmatrix an. Dabei ist auch hier auf eine Einbeziehung der Altersstruktur der Daten zu achten. Weil im Fall verschiedener gradueller Änderungen diese nur unter Vernachlässigung der Zugehörigkeitsgrade deutlich werden können, wird auch in der Berechnung des Gewichtungsparameters γ_i^{Σ} auf diese verzichtet, so dass sich analog zu (5.4)

$$\gamma_i^{\Sigma} = \frac{\gamma^{\Sigma} n_i^{\alpha^{\mathrm{A}neu}}}{(\gamma^{\Sigma} - 1)\, n_i^{\alpha^{\mathrm{A}neu}} + n_i^{\alpha^{\mathrm{A}}}} \tag{5.12}$$

ergibt; dabei ist γ^{Σ} ein allgemeiner Gewichtungsparameter für den Einfluss neuer Objekte[20]. Analog zu (5.4) ist die Wahl von γ^{Σ} auch hier kontextabhängig, im Idealfall gilt $\gamma^{\vec{v}} = \gamma^{\Sigma}$ für konsistente Ergebnisse.

Anhand der neuen Kovarianzmatrizen lassen sich bereits Hinweise auf eine Veränderung der Ausdehnung eines Clusters innerhalb der einzelnen Dimensionen anhand der Veränderungen der Varianzen ableiten. Nimmt eine Varianz zu, so dehnt sich das Cluster weiter aus; es liegt

[20] Alternativ ist auch die Gewichtung mit $\gamma_i^{\vec{v}}$ möglich. Dies beschleunigt die Berechnung, da kein neuer Parameter bestimmt werden muss, führt jedoch gleichzeitig zu einer Einbeziehung der Zugehörigkeitsgrade, die bei Schätzung der harten Kovarianzmatrix nicht sinnvoll ist. Empirisch hat die unterschiedliche Bestimmung der Gewichtung jedoch nur sehr geringe Auswirkungen, weshalb sie aus Gründen des Rechenaufwands je nach Anwendungsfall in Kauf genommen werden kann.

also eine Volumenerhöhung vor. Nimmt sie hingegen ab, reduziert sich das Clustervolumen. Insbesondere durch Anwendung des Fuzzy-Hypervolumens $FHV\left([U_i]_{\alpha^{\mathrm{A}}}\right)$ aus (5.13) (vgl. z.B. Höppner u. a., 1999, S. 193) bezogen auf den α-Schnitt, das die allgemeine Ausdehnung eines Clusters beschreibt, können Veränderungen im Clustervolumen aufgedeckt werden. Unabhängig von seinem Namen kann das Fuzzy-Hypervolumen auch auf harte Kovarianzen angewandt werden.

$$FHV\left([U_i]_{\alpha^{\mathrm{A}}}\right) = \sqrt[p]{\det\left(\Sigma_i^{hart}\right)} \tag{5.13}$$

Ferner geben die Eigenwerte θ_{il} der Kovarianzmatrix Σ_i^{hart} weitere Informationen bzgl. der Clusterausdehnung. Außerdem kann anhand der Korrelationen sowie der Ausrichtung der Eigenvektoren der Kovarianzmatrix $\vec{e}_{il}$, $l = 1, \ldots, p$, bzw. anhand ihres Skalarprodukts zu verschiedenen Zeitpunkten, $(|\vec{e}_{il}^{\,t-\Delta t} \cdot \vec{e}_{il}^{\,t}|)$, aufgedeckt werden, dass sich die generelle Clusterausrichtung ändert, das Cluster sich also dreht. Ist der Wert nahezu Eins, liegt keine signifikante Drehung vor; je kleiner das Skalarprodukt der Eigenvektoren ist, desto stärker wird das Cluster rotiert. Hierbei ist eine korrekte Zuordnung zu beachten, da sich unter Umständen nicht die Ausrichtung allgemein, sondern lediglich die Ausdehnung innerhalb der Dimensionen geändert hat, so dass die Reihenfolge der Eigenwerte und der zugehörigen Eigenvektoren von der vorherigen Analyse abweicht.

Tabelle 5.7 stellt die eingeführten Parameter den möglichen Veränderungen gegenüber (vgl. Minke und Ambrosi, 2012).

Veränderung der Objektzahl

Wandelt sich die Anzahl an Objekten innerhalb eines Clusters, wird diese Änderung durch die absolute sowie die Fuzzy-Kardinalität deutlich. Sinkt die Objektzahl zwischen zwei Analysezeitpunkten, so verringern sich auch die Kardinalitäten; steigt sie, nehmen die Kardinalitäten zu. Weder der Kompaktheitsindex noch der lokale Partitionskoeffizient zeigen signifikante Änderungen. Abhängig von der Kovarianzmatrix im Falle ellipsoider Cluster können bzgl. der Clusterausdehnung leichte Schwankungen anhand des Fuzzy-Hypervolumens und der Eigenwerte festgestellt werden. Diese sind jedoch unabhängig von der Richtung der aufgetretenen Veränderung, sondern hängen davon ab, wo innerhalb des Clusters Objekte hinzugekommen bzw. weggefallen sind. Das Skalarprodukt der Eigenvektoren ist ungefähr Eins, da keine signifikante Änderung der Clusterausrichtung vorliegt.

Veränderung des Clustervolumens

Im Folgenden wird eine über die Dimensionen hinweg gleichmäßige Volumenänderung innerhalb eines Clusters angenommen. Wenn das Volumen eines Clusters kontinuierlich abnimmt, d.h., wenn die Objekte sich einander annähern, ergibt diese Entwicklung Konsequenzen für verschiedene Maße. Die Volumenreduktion bewirkt eine leichte Erhöhung der absoluten sowie der Fuzzy-Kardinalität, da aufgrund der Annäherung und der damit einhergehenden höheren Zugehörigkeitsgrade neuer Objekte mehr Objekte durch das Cluster absorbiert werden. Der Kompaktheitsindex sinkt und verdeutlicht dadurch ein Erhöhung der Kompaktheit des Clusters. Aus demselben Grund steigt der lokale Partitionskoeffizient an. Im Zusammenhang mit

Parameter	Veränderung Objektzahl	Veränderung Volumen	Veränderung Ausrichtung
$n_i^{\alpha^{\mathrm{A}}}$	Objektzahlreduktion: deutlicher Rückgang Objektzahlerhöhung: deutliche Zunahme	Volumenerhöhung: leichter Rückgang Volumenreduktion: leichte Zunahme	leichter Rückgang
$\mathrm{card}\left([U_i]_{\alpha^{\mathrm{A}}}\right)$	Objektzahlreduktion: deutlicher Rückgang Objektzahlerhöhung: deutliche Zunahme	Volumenerhöhung: Rückgang Volumenreduktion: Zunahme	keine signifikanten Veränderungen
$\kappa_i^{\alpha^{\mathrm{A}}}$	keine signifikanten Veränderungen	Volumenerhöhung: Zunahme Volumenreduktion: Rückgang	nahezu konstant, leichte Schwankungen aufgrund Einbeziehung alter Kovarianzmatrix
$LPC\left([U_i]_{\alpha^{\mathrm{A}}}\right)$	keine signifikanten Veränderungen	Volumenerhöhung: Rückgang Volumenreduktion: Zunahme	keine signifikanten Veränderungen
$FHV\left([U_i]_{\alpha^{\mathrm{A}}}\right)$	leichte Schwankungen unabhängig von Richtung der Veränderung	Volumenerhöhung: Zunahme Volumenreduktion: Rückgang	keine signifikanten Veränderungen
θ_{il}	leichte Schwankungen unabhängig von Richtung der Veränderung	Volumenerhöhung: Zunahme Volumenreduktion: Rückgang	leichte Schwankungen
$\lvert\vec{e}_{il}^{\,t-\Delta t} \cdot \vec{e}_{il}^{\,t}\rvert$	≈ 1 (keine Änderung)	≈ 1 (keine Änderung)	< 1 (Änderung der Ausrichtung)

Tabelle 5.7.: Zuordnung Parameter zur Messung gradueller Veränderungen

einer Volumenerhöhung sind die Auswirkungen auf die lokalen Strukturmaße weniger deutlich; dieser Umstand wird durch den durch α^{A} begrenzten Betrachtungsraum begründet. Liegt eine Volumenerhöhung vor, verlassen Objekte den Betrachtungsraum, da ihre Zugehörigkeitsgrade auf Basis der alten Clusterzentren und damit auf Basis des alten Volumens geschätzt werden. Als Folge zeigen sich die Auswirkungen insbesondere auf den Kompaktheitsindex und den lokalen Partitionskoeffizienten deutlich geringer. Dieser Umstand stellt jedoch nur bei runden Clustern ein Problem dar. Wird außerdem die Kovarianzmatrix betrachtet, zeigen das Fuzzy-Hypervolumen und die Änderung der Eigenwerte die Veränderung an. Das Skalarprodukt der Eigenvektoren ist nahezu Eins, da die Clusterausrichtung unverändert bleibt.

Veränderung der Ausrichtung

Die Clusterausrichtung kann sich nur bei ellipsoiden Clustern ändern, da nur dort die einzelnen Dimensionen einer unterschiedlichen Entwicklung unterliegen können. Ändert sich nur die Ausrichtung, während alle anderen strukturellen Charakteristika konstant bleiben, zeigen die lokalen Maße keine signifikanten Änderungen an. Lediglich die absolute Kardinalität kann leicht sinken, da Objekte durch die Clusterdrehung den Betrachtungsraum verlassen; analog lassen sich leichte Schwankungen in den Eigenwerten begründen. Anhand der Kovarianzmatrix des Clusters und den zugehörigen Eigenvektoren wird die Änderung deutlich: Aufgrund der Rotation des Clusters können signifikante Änderungen der Eigenvektoren verzeichnet werden; dargestellt wird diese Entwicklung durch das Skalarprodukt der Eigenvektoren zweier aufeinanderfolgender Analysezeitpunkte von deutlich unter Eins.

Kombination verschiedener struktureller Änderungen

I.d.R. treten die einzelnen Veränderungen eines Clusters nicht isoliert auf; vielmehr gilt es, eine Kombination verschiedener Veränderungen zu identifizieren. Tritt eine Clusterdrehung auf, so kann diese unabhängig von weiteren Veränderungen aufgedeckt werden, da sie als einzige anhand des Skalarproduktes der Eigenvektoren zu unterschiedlichen Zeitpunkten erkennbar wird und außerdem das Skalarprodukt zur Identifikation der Rotation ausreicht. Tritt jedoch eine Veränderung der Objektzahl gleichzeitig mit einer Volumenänderung auf, wird die Analyse erschwert. Die Auswirkungen auf die Parameter ergeben sich aus der Kombination der entsprechenden Spalten in Tabelle 5.7. Es ist jedoch nicht möglich, allgemeingültige Aussagen zum konkreten Ausmaß der Auswirkungen zu machen, da die Auswirkungen von der Stärke der einzelnen Veränderungen sowie dem Verhältnis der unterschiedlichen Anpassungsarten zueinander abhängen. Im Extremfall können sich einzelne, gegenläufige Veränderungen sogar aufheben. Um ein gemeinsames Auftreten dieser Veränderungen zu identifizieren, wird daher zunächst die deutlichere Veränderung angenommen. Die dadurch verursachten Abweichungen werden gemäß Tabelle 5.7 eliminiert. Sollte eine erwartete Parameterauswirkung nicht vorhanden sein, wird diese nachfolgend in die entgegengesetzte Richtung verschoben, da die Vermutung naheliegt, dass sich die stattfindenden Veränderungen gegenseitig kompensiert haben. Anschließend kann evaluiert werden, inwiefern eine weitere Veränderung vorliegt.

Wie anfänglich beschrieben, stellt die Veränderung der Verteilung innerhalb eines Clusters und die damit einhergehende Implikation einer bevorstehenden Clustertrennung einen Spezialfall der graduellen Unterschiede dar. In diesem Fall sind weiterführende Untersuchungen

vorzunehmen, die bei einer Evaluation rein gradueller Änderungen jedoch von untergeordneter Bedeutung sind. Da diese Art der Änderung lediglich einen Spezialfall für eine bevorstehende Clustertrennung darstellt, wird sie im dazugehörigen Abschnitt 5.7 detailliert betrachtet.

Anhand der gemessenen graduellen Unterschiede erfolgt im Anschluss eine Zuordnung der möglichen abrupten Veränderungen bei graduellen Unterschieden in der Clusterstruktur (vgl. Tabelle 5.2) gemäß Tabelle 5.8. Die konkrete Analyse der abrupten Veränderung zur Clusterelimination bzw. zur Clustertrennung wird im jeweiligen Abschnitt detailliert aufgezeigt.

Abrupte Veränderung	**impliziert durch**
Clusterelimination	Reduktion der Objektzahl oder des Clustervolumens
Clustertrennung	Verwaisung des Clusterzentrums

Tabelle 5.8.: Zuordnung konkreter gradueller Veränderungen

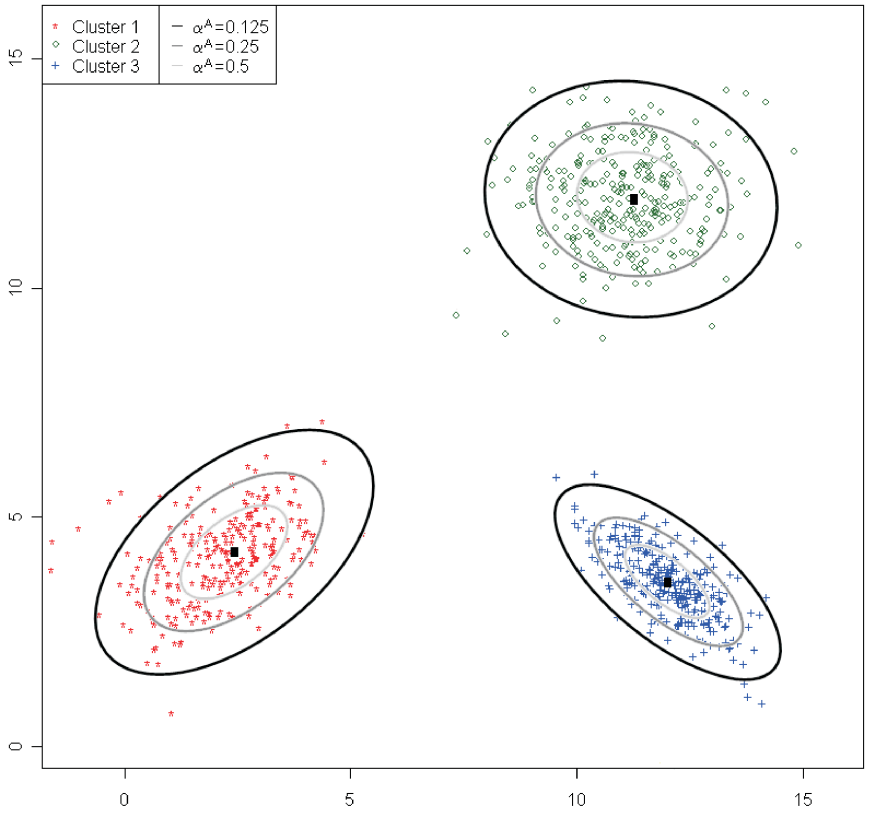
Analog zu den Änderungen bzgl. der Clusterpositionierung müssen auch die Veränderungen der Fuzzy-Kardinalität und des Kompaktheitsindexes in der Clusterhistorie abgelegt werden, so dass Trends identifiziert und mögliche abrupte Veränderungen prognostiziert werden können.

Im Folgenden soll das Vorgehen anhand eines Beispiels veranschaulicht werden. Es wurde ein zweidimensionaler Datensatz mit drei elliptischen Clustern normalverteilter Daten kreiert. Die konkreten Parameter sind in Tabelle 5.9 gegeben. Während die Parameter für Cluster 2 und Cluster 3 unverändert blieben, wurden im ersten Cluster mehrere strukturelle Änderungen vorgenommen. Ausgehend von 100 Objekten je Cluster wurde die Anzahl neu hinzukommender Objekte je Periode um zehn Objekte verringert, ferner wurde das allgemeine Volumen sukzessive erhöht und eine Drehung um 10° je Periode durchgeführt. Analog zum Beispiel zur Clusterverschiebung (vgl. Abschnitt 5.3.1) wurden eine Zeitfensterlänge von $\tau = 3$ sowie eine Analysefrequenz von $\Delta t = 1$ gewählt. Insgesamt wurden vier Perioden betrachtet, wobei die Analysen in der dritten und in der vierten Periode stattfanden. Als Gewichtungsparameter beim inkrementellen Update wurde $\gamma^{\vec{v}} = \gamma^{\Sigma} = \tau$ gewählt. Der Grenzwert der Absorbierung wurde auf $\alpha^{\mathrm{A}} = 0.125$ festgelegt.

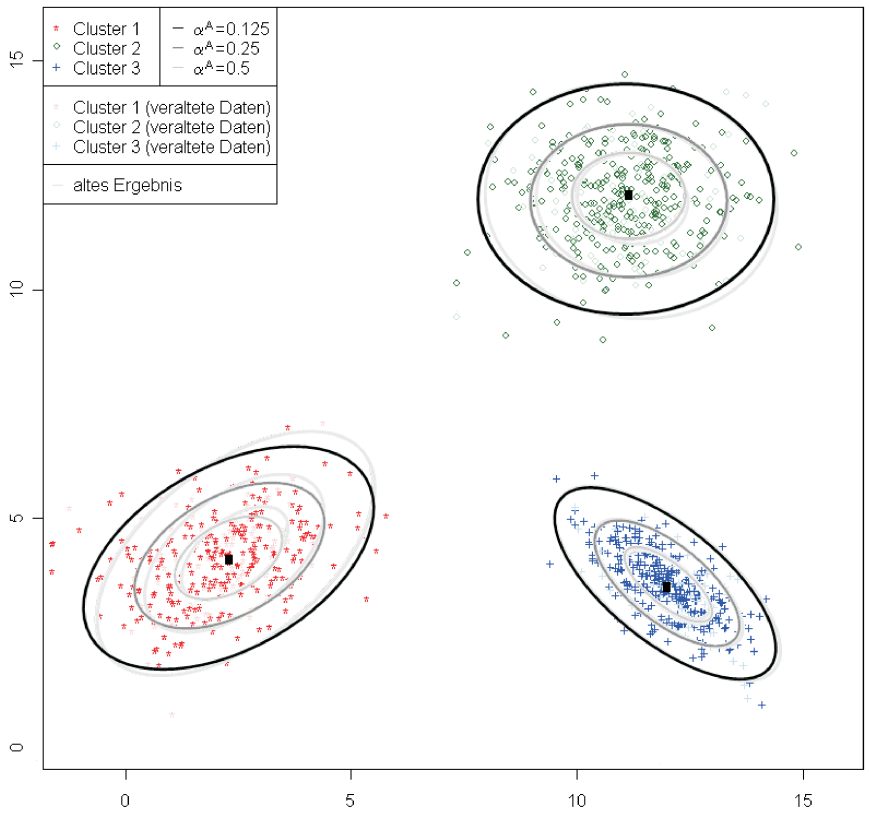
In Abbildung 5.6 sind die Ergebnisse des Gustafson-Kessel-Algorithmus unter Einbeziehung der Dreiecksbeziehung der Distanzen dargestellt (vgl. Abschnitt 4.3.3, Algorithmus 4.7). Abbildung 5.6a veranschaulicht die Clusterstruktur zum Zeitpunkt der ersten Analyse in Periode $t = 3$. Darauf aufbauend erfolgt eine Schätzung der Veränderungen, die durch die in Periode

Cluster	**Zentrum**	**Kovarianzmatrix**
Cluster 1	$\begin{pmatrix} 2 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$
Cluster 2	$\begin{pmatrix} 11 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 1.5 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$
Cluster 3	$\begin{pmatrix} 12 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 0.8 & -0.5 \\ -0.5 & 0.6 \end{pmatrix}$

Tabelle 5.9.: Vorgegebene Parameter in erster Periode

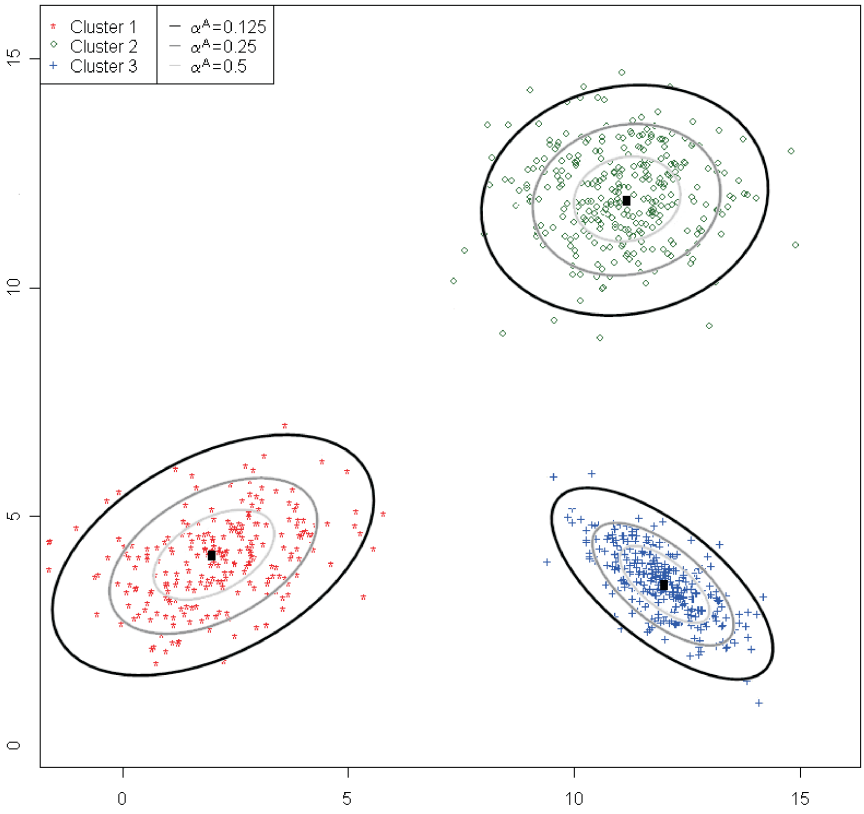


(a) Clusterstruktur für den Zeitpunkt $t = 3$ (Perioden 1-3)



(b) Schätzung der neuen Clusterstruktur für $t = 4$ für neue Daten aus (Perioden 2-4)

Abbildung 5.6.: Beispiel für graduelle Änderungen innerhalb eines Clusters



(c) Clusterstruktur für das Zeitpunkt $t = 4$ (Perioden 2-4)

Abbildung 5.6.: Beispiel für graduelle Änderungen innerhalb eines Clusters

$t = 4$ neu hinzugekommenen Objekte angezeigt werden (Abbildung 5.6b). Die verschiedenen Parameter zur Evaluation gradueller Veränderungen werden in Tabelle 5.10 gegenübergestellt. Während für Cluster 2 und Cluster 3 nur geringfügige, weitestgehend nicht signifikante Abweichungen innerhalb der Daten nachgewiesen werden konnten, fallen für das erste Cluster starke Änderungen auf. Auffällig ist die Änderung der volumenbezogenen Maße, die eine deutliche Erhöhung des Volumens anzeigen (vgl. Tabelle 5.7). Der gravierende Rückgang innerhalb der Kardinalitäten kann nur teilweise mit dem steigenden Clustervolumen begründet werden, d.h., dass eine weitere Ursache vorliegen muss. Auf diese Weise wird die sinkende Objektzahl innerhalb des Clusters angezeigt. Die Änderung der Clusterausrichtung wird durch die Eigenvektoren und ihre Skalarprodukte sichtbar, die jeweils eine Drehung der Vektoren um 8,2° zeigen.

Zur Vergleichbarkeit ist in Abbildung 5.6c das Ergebnis einer separaten Analyse zu diesem Zeitpunkt dargestellt. Durch die Schätzung auf Basis der neuen Daten in Abbildung 5.6b erfolgt eine gute Annäherung an das Gesamtergebnis.

Cluster	**Parameter**	**Analyseergebnis für** $t = 3$	**Schätzung für** $t = 4$
Cluster 1	$n_1^{\alpha^A}$	258	220
	$\text{card}\left([U_1]_{\alpha^A}\right)$	123.0605	94.9164
	$\kappa_1^{\alpha^A}$	0.4108	0.4506
	$LPC\left([U_1]_{\alpha^A}\right)$	0.2820	0.2380
	$FHV\left([U_1]_{\alpha^A}\right)$	0.8326	1.0184
	θ_{11}	1.5415	1.8046
	θ_{12}	0.4496	0.5747
	$\vec{e}_{11}$	$\begin{pmatrix} -0.7938 \\ 0.6081 \end{pmatrix}$	$\begin{pmatrix} -0.8724 \\ 0.4887 \end{pmatrix}$
	$\lvert\vec{e}_{11}^{\,3} \cdot \vec{e}_{11}^{\,4}\rvert$		0.9898
	$\vec{e}_{12}$	$\begin{pmatrix} 0.6081 \\ 0.7938 \end{pmatrix}$	$\begin{pmatrix} 0.4887 \\ 0.8724 \end{pmatrix}$
	$\lvert\vec{e}_{12}^{\,3} \cdot \vec{e}_{12}^{\,4}\rvert$		0.9898
Cluster 2	$n_2^{\alpha^A}$	285	286
	$\text{card}\left([U_2]_{\alpha^A}\right)$	131.3697	133.9199
	$\kappa_2^{\alpha^A}$	0.5082	0.5018
	$LPC\left([U_2]_{\alpha^A}\right)$	0.2645	0.2712
	$FHV\left([U_2]_{\alpha^A}\right)$	1.0769	1.0929
	θ_{21}	1.3820	1.4256
	θ_{22}	0.8391	0.8378
	$\vec{e}_{21}$	$\begin{pmatrix} -0.9972 \\ -0.0747 \end{pmatrix}$	$\begin{pmatrix} -1.0000 \\ -0.0097 \end{pmatrix}$
	$\lvert\vec{e}_{21}^{\,3} \cdot \vec{e}_{21}^{\,4}\rvert$		0.9964
	$\vec{e}_{22}$	$\begin{pmatrix} -0.0747 \\ 0.9972 \end{pmatrix}$	$\begin{pmatrix} 0.0097 \\ 1.0000 \end{pmatrix}$
	$\lvert\vec{e}_{22}^{\,3} \cdot \vec{e}_{22}^{\,4}\rvert$		0.9964
Cluster 3	$n_3^{\alpha^A}$	292	292
	$\text{card}\left([U_3]_{\alpha^A}\right)$	147.2899	147.7079
	$\kappa_3^{\alpha^A}$	0.2107	0.2118
	$LPC\left([U_3]_{\alpha^A}\right)$	0.3117	0.3115
	$FHV\left([U_3]_{\alpha^A}\right)$	0.4263	0.4326
	θ_{31}	1.0247	1.0067
	θ_{32}	0.1773	0.1859
	$\vec{e}_{31}$	$\begin{pmatrix} -0.7805 \\ -0.6251 \end{pmatrix}$	$\begin{pmatrix} -0.7916 \\ -0.6111 \end{pmatrix}$
	$\lvert\vec{e}_{31}^{\,3} \cdot \vec{e}_{31}^{\,4}\rvert$		0.9998
	$\vec{e}_{32}$	$\begin{pmatrix} -0.6251 \\ 0.7805 \end{pmatrix}$	$\begin{pmatrix} -0.6111 \\ 0.7916 \end{pmatrix}$
	$\lvert\vec{e}_{32}^{\,3} \cdot \vec{e}_{32}^{\,4}\rvert$		0.9998

Tabelle 5.10.: Lokale Strukturveränderungen

5.4. Neubildung von Clustern

Im Zeitablauf ist es möglich, dass neben bereits existierenden Clustern neue entstehen. Dem Erkennen dieser neugebildeten Cluster kommt beim Change Mining eine hohe Bedeutung zu, da nur durch Aufdecken der Veränderung die Clusterstruktur aktuell gehalten werden kann. Im Bereich der Marktforschung und des Marketings ist dieser Umstand von besonderer Relevanz, da sich eine zeitgemäße Marktabbildung für den sinnvollen Einsatz der verschiedenen Marketing-Maßnahmen als elementar erweist (vgl. Abschnitt 1.4). Werden aus Unkenntnis über ihr Vorhandensein neue Cluster – und damit neue Segmente – vernachlässigt, wirkt sich dies für das betreffende Unternehmen unter Umständen nachteilig aus. Entsteht z.B. ein neues Kundensegment, dessen differenzierte Bearbeitung sich als gewinnbringend für das Unternehmen erweisen könnte, dieses aber erst nach Ausbreitung anderer Anbieter in dem betroffenen Segment reagiert, kann ein nicht unerheblicher Schaden entstehen. Auch das Aufdecken von Marktnischen erhält in diesem Zusammenhang eine hohe Bedeutung, da ihre Bearbeitung nur dann potentiell ratsam erscheint, solange hier noch kein (Über-)Angebot vorherrscht.

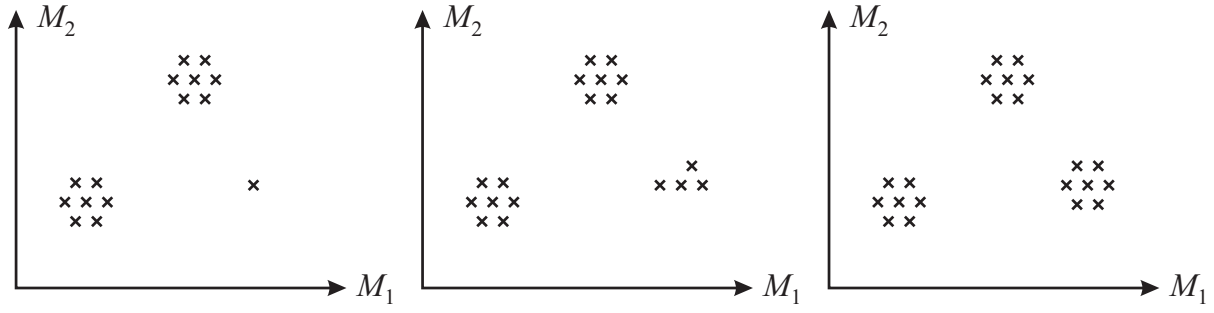
Neben dem einfachen Erkennen neuer Cluster zählt auch die Prognose möglicher entstehender Cluster zu den bedeutenden Faktoren, da sie es einem Unternehmen ermöglicht, frühzeitig auf diese Veränderung zu reagieren.

Im Folgenden wird zunächst das Vorgehen zum Erkennen und Prognostizieren der Neubildung von Clustern eingeführt, bevor dieses Vorgehen anhand eines künstlichen Beispiels veranschaulicht wird.

5.4.1. Vorgehen zum Erkennen der Neubildung

Ein neues Cluster entsteht i.d.R. nicht ad hoc; vielmehr ist eine Zunahme der Objektzahl innerhalb einer Region zu verzeichnen; Abbildung 5.7 soll diesen Umstand verdeutlichen. Zum Zeitpunkt t_1 sind zunächst nur zwei gut separierte Cluster zu erkennen (Abbildung 5.7a). Das zusätzlich auftretende Objekt im unteren rechten Bildabschnitt wird als Ausreißer erkannt (vgl. Definition 5.1), scheint zu diesem Zeitpunkt jedoch noch keinerlei Bedeutung zu besitzen. Zum darauffolgenden Zeitpunkt t_2 sind die beiden zuvor erkannten Cluster unverändert (Abbildung 5.7b). Zu dem als Ausreißer ermittelten Objekt sind jedoch weitere Objekte hinzugekommen. Damit ist theoretisch bereits ein drittes Cluster zu erkennen; es bleibt jedoch zu prüfen, ob diese Anhäufung an Objekten bereits eine ausreichende Größe vorweisen kann, um als eigenständiges Cluster behandelt zu werden. Reicht diese nicht aus, darf die Zunahme der Objektzahl in dieser Region jedoch nicht wieder verworfen werden: Sie ist ein Indiz für das Entstehen eines Clusters in diesem Gebiet. In der dritten Abbildung 5.7c, die den Zeitpunkt t_3 wiedergibt, besitzt das neue Cluster dieselbe Größe wie die beiden zum Zeitpunkt t_1 ermittelten und ist demnach ebenso zu berücksichtigen.

In der Realität kann die Bestimmung neu entsehender Cluster nicht derart intuitiv erfolgen. Der Grund hierfür liegt zum einen in den höherdimensionierten Räumen, da i.d.R. keine Analysen im zweidimensionalen Raum, sondern im $\mathbb{R}^p$ durchgeführt werden, zum anderen in der nicht wie im konstruierten Fall derartig eindeutig vorhandenen Struktur. Die übrigen Cluster



(a) Zeitpunkt t_1 (b) Zeitpunkt t_2 (c) Zeitpunkt t_3

Abbildung 5.7.: Entstehung eines neuen Clusters

verändern sich, außerdem unterliegen Ausreißer – wie ihr Name bereits ausdrückt – meist keiner eindeutigen Struktur, vielmehr sind sie nahezu beliebig im Raum verteilt. Dennoch sollen Veränderungen dieser Art in einem höherdimensionierten Datensatz sichtbar gemacht werden. Dies erfolgt analog zur intuitiven Bestimmung in dem vereinfachten Beispiel aus Abbildung 5.7 in zwei wesentlichen Schritten:

1. Identifikation von Ausreißern
2. Bestimmung dichter Regionen innerhalb der Ausreißer

Zur Identifikation von Ausreißern werden die Zugehörigkeitsgrade der einzelnen Objekte herangezogen. Objekte mit einem geringen Zugehörigkeitsgrad zu allen Clustern, d.h. alle Objekte, deren Zugehörigkeitsgrade den vorgegebenen Grenzwert α^{A} unterschreiten (vgl. Definition 5.1), sind als Ausreißer bzgl. der aktuellen Clusterstruktur zu betrachten.

Definition 5.1. Sei ein Objekt j gegeben durch seinen Eigenschaftsvektor $\vec{x}_j$. Das Objekt heißt *Ausreißer* bzgl. einer Clusterstruktur $C = \{C_1, \ldots, C_c\}$, falls für die Zugehörigkeitsgrade u_{ij} gilt

$$u_{ij} < \alpha^{\mathrm{A}} \; \forall i = 1, \ldots, c,$$

wobei α^{A} den Grenzwert der Absorbierung angibt.

Als Grenzwert der Absorbierung sollte ein kleiner Wert gewählt werden (vgl. Abschnitt 5.3). Bei der Wahl großer Werte für α^{A} werden viele Objekte fälschlicherweise als echte Ausreißer identifiziert, bei kleinen Werten hingegen kann sichergestellt werden, dass die identifizierten Ausreißer tatsächlich als solche definiert werden können. Ferner ist die Wahrscheinlichkeit, ein potentiell neues Cluster im direkten Umfeld eines bereits bestehenden zu finden, bei kleinen α^{A}-Werten deutlich geringer als bei größeren.

Die bloße Existenz von Ausreißern reicht zur Entstehung neuer Cluster nicht aus, wie in Abbildung 5.7a verdeutlicht ist. Erst wenn ihre Anzahl einen vorgegebenen Grenzwert überschreitet, kann das Vorhandensein von neuen Clustern innerhalb der Ausreißer in Erwägung gezogen werden. Angstenberger (2000, S. 85f.) schlägt die Wahl eines Grenzwertes vor, so dass die Mindestgröße eines neu entstehenden Clusters abhängig von der Größe des kleinsten Clusters nach der Absorbierung mit Hilfe des Grenzwerts α^{A} ist:

$$n^{\mathrm{A}} \geq \beta_{nC}^{\min} \cdot n_{\min}^{\alpha^{\mathrm{A}}}, \tag{5.14}$$

wobei

- n^{A} – Anzahl Ausreißer nach Definition 5.1,
- $n_{\min}^{\alpha^{\mathrm{A}}}$ – Anzahl an Objekten, die durch den α-Schnitt mit α^{A} durch das kleinste Cluster absorbiert werden, d.h. $n_{\min}^{\alpha^{\mathrm{A}}} = \min\left\{n_i^{\alpha^{\mathrm{A}}} | i \in \{1,\ldots,c\}\right\}$,
- $\beta_{nC}^{\min}$ – Parameter zur Bestimmung des geforderten Anteils.

Um eine zukünftige Entwicklung frühzeitig aufdecken zu können, empfiehlt es sich, den Parameter $\beta_{nC}^{\min}$ eher klein zu wählen; ein zu großer Wert kann dazu führen, dass bei Strukturen wie in Abbildung 5.7b ein entstehendes Cluster nicht erkannt wird. Eine frühzeitige Identifikation potentiell wachsender Cluster bietet zudem den Vorteil, dass ggf. andere, bereits existierende Cluster bei der Analyse mit einer zu kleinen Clusterzahl nicht durch Häufungen von Ausreißern bzgl. Lage und Ausrichtung beeinflusst werden.

Bei einer ausreichend vorhandenen Anzahl an Ausreißern werden innerhalb dieser dichte Regionen ermittelt, sogenannte *Ausreißercluster*. Diese Ausreißercluster sind im dynamischen Kontext zu betrachten, da sowohl ihr Entstehungszeitpunkt als auch ihr Wachstum Aufschluss darüber geben, ob an dieser Stelle ein relevantes Cluster entsteht (vgl. Definition 5.2).

Definition 5.2. (Vgl. Cao u. a., 2006) Ein Ausreißercluster zum Zeitpunkt t für eine Gruppe naher Objekte $\vec{x}_{t_1}^{\mathrm{A}},\ldots,\vec{x}_{t_j}^{\mathrm{A}},\ldots,\vec{x}_{t_{n^{\mathrm{A}}}}^{\mathrm{A}}$ ist definiert als $C_{t_i}^{\mathrm{A}} = \left(\vec{v}_{t_i}, \eta_{t_i}(,\Sigma_{t_i}), t_i^0, \Delta n_{t_i}^{\alpha^{\mathrm{A}}}, \Delta PD_{t_i}\right)$, wobei

- t_i^0 – Zeitpunkt des erstmaligen Auftretens des Ausreißerclusters i,
- $\Delta n_{t_i}^{\alpha^{\mathrm{A}}}$ – Veränderung der Clustergröße gegenüber der Vorperiode; falls $t = t_i^0$: $\Delta n_{t_i}^{\alpha^{\mathrm{A}}} := n_{t_i}^{\alpha^{\mathrm{A}}}$,
- ΔPD_{t_i} – Veränderung der Clusterdichte gegenüber der Vorperiode; falls $t = t_i^0$: $\Delta PD_{t_i} := PD_{t_i}$.

Die Ausreißercluster werden auf Basis einer Clusteranalyse ermittelt. Dabei werden nur die Ausreißer in die Analyse einbezogen, durch vorhandene Cluster absorbierte Objekte werden nicht beachtet. Die gefundene Anzahl an Ausreißern liefert die Obergrenze potentieller neuer Cluster gemäß (5.14) mit

$$c_t^{pot} = \left\lfloor \frac{n^{\mathrm{A}}}{\beta_{nC}^{\min} \cdot n_{\min}^{\alpha^{\mathrm{A}}}} \right\rfloor . \tag{5.15}$$

Zur Ermittlung der geeigneten Clusterzahl kann bei Bedarf ein hierarchisches Verfahren wie z.B. Ward's Linkage hinzugezogen werden (vgl. z.B. Backhaus u. a., 2006, S. 517ff.). Durch die anschließende Durchführung einer einfachen possibilistischen Clusteranalyse (vgl. Abschnitt 3.4) gelingt es, dichte Regionen zu ermitteln. Dabei werden die clusterspezifischen Parameter η_i einmalig neu bestimmt (vgl. Abschnitt 3.4). Zöge man lediglich die aus der probabilistischen Analyse ermittelten Werte für die η_i hinzu, besäße der entsprechende Wert insbesondere für den Spezialfall $c_t^{pot} = 1$ keinerlei Aussagekraft, da in diesem Fall für alle probabilistischen Zugehörigkeitsgrade aufgrund der wahrscheinlichkeitsbasierten Nebenbedingung in (3.3) $u_{ij} = 1$ gelten würde. Auch für $c_t^{pot} > 1$ kann sich diese Bedingung als problematisch erweisen, da nicht zwischen solchen Objekten unterschieden wird, die zu Ausreißerclustern gehören, und solchen, die echte Ausreißer sind.

Die Clusteranziehung der klassischen possibilistischen Analyse stellt in diesem Zusammenhang ein vernachlässigbares Problem dar, da gut separierte Häufungspunkte innerhalb der Ausreißer gesucht werden; sie kann sich sogar als vorteilhaft erweisen, falls die verwendete Clusterzahl zu hoch gewählt wurde. Aufgrund der Clusteranziehung finden Cluster mit zu geringer Dichte oder in direkter Nähe anderer Cluster keine Berücksichtigung bei Anwendung der einfachen possibilistischen Analyse; vielmehr werden andere Cluster mehrfach identifiziert. Um die tatsächliche Anzahl neuer Cluster zu bestimmen, ist es erfolderlich, in diesem Fall die mehrfach vorhandenen Cluster gesondert zusammenzufassen. Dies geschieht weitestgehend analog zur Clustervereinigung (vgl. Abschnitt 5.6), beispielsweise durch eine einfache Verwendung der Ähnlichkeit zwischen Clustern auf Basis des Inklusionsmaßes in (5.38).

Alternativ zur Verwendung der possibilistischen Analyse kann die Bestimmung der tatsächlichen Clusterzahl innerhalb der Ausreißer auch mit Hilfe des von Li und Mukaidono (1995) vorgestellten Ansatzes erfolgen: Die Autoren schlagen vor, zu verschiedenen Clusterzahlen $c_t \in \left[c_t^{pot_{\min}}, c_t^{pot_{\max}}\right]$ eine probabilistische Analyse durchzuführen und das Ergebnis im Anschluss bzgl. seiner *Strukturstärke* zu evaluieren. Hierzu wird eine Schadensfunktion $L(c_t)$ verwendet, die eine Fuzzy-Variante der Innergruppenvarianz darstellt:

$$L(c_t) = \sum_{i=1}^{c_t} \sum_{j=1}^{n^{\mathrm{A}}} u_{ij} d^2_{C_i}(\vec{v}_i, \vec{x}_j) \tag{5.16}$$

Basierend auf der Innergruppenvarianz wird im Anschluss unter Einbeziehung der Effektivität der Clusterzuordnung und ihrer Genauigkeit die Strukturstärke $St(c_t)$ in (5.17) bei einer gegebenen Clusterzahl bestimmt:

$$St(c_t) = \gamma^{St} \ln \frac{n^{\mathrm{A}}}{c_t} + (1 - \gamma^{St}) \ln \frac{L(1)}{L(c_t)}, \tag{5.17}$$

wobei γ^{St}ein Gewichtungsparameter zur Gewichtung der beiden Terme ist; Li und Mukaidono (1995) empfehlen eine ungewichtete Schätzung, d.h. $\gamma^{St} = 0.5$.

Der erste Term in (5.17) bewertet diejenigen Strukturen als effektiver und damit als besser, die nur wenige Cluster benötigen. Bis zu einem gewissen Grad erscheint dieses Vorgehen sinnvoll, da im Rahmen der Clusteranalyse wenige homogene Cluster aufgedeckt werden sollen. Der zweite Term bewertet hingegen diejenigen Strukturen als genauer, die viele Cluster verwenden. Auch dieser Ansatzpunkt erweist sich als relevant, da die genaueste Repräsenation, d.h. die Repräsentation mit dem geringsten Informationsverlust, dann erreicht wird, wenn jedes Objekt als eigenes Cluster dargestellt wird; dabei ist der Extremfall $c_t = n^{\mathrm{A}}$ zwar der genaueste, jedoch der am wenigsten brauchbare, weil keine echte Clusterstruktur mehr ermittelt wird. Da die Ziele sich gegenseitig ausschließen, jedoch zu einem gewissen Grad von Bedeutung sind, erfolgt eine (ggf. über γ^{St} gewichtete) Bestimmung der Clusterstärke anhand der Teilziele. Die geeigneteste Clusterzahl lässt sich durch Maximierung der Clusterstärke bzgl. verschiedener Clusterzahlen c_t^{pot} bestimmen.

Die Bestimmung der Clusterzahl auf Basis der Strukturstärke ist zwar genauer als eine grobe Schätzung der Clusterzahl mit anschließender possibilistischer Analyse, sie ist jedoch auch um einiges aufwändiger, da für verschiedene Clusterzahlen eine probabilistische Clusteranalyse durchgeführt werden muss. Aufgrund dieser Tatsache scheint es sinnvoller, auf die genauere

Analyse an dieser Stelle zu verzichten. Bei Bedarf kann sie jedoch am Ende des Change Mining-Prozesses zur Bestimmung der neuen Struktur herangezogen werden, um die durch die Analyse der verschiedenen abrupten Veränderungen angepasste Clusterzahl zu verifizieren.

Sind die dichten Regionen ermittelt, bedeutet dies noch nicht, dass an der jeweiligen Stelle tatsächlich ein neues Cluster entsteht bzw. bereits entstanden ist. Hierzu müssen Größe und Dichte des Clusters näher untersucht sowie sein direktes Umfeld in der Gesamtstruktur der Daten analysiert werden. Zunächst muss sichergestellt sein, dass ein potentiell neues Cluster sich nicht in der direkten Umgebung eines bereits bestehenden Clusters befindet. Bei kleinen Werten für α^{A} kann dies bei vorheriger Untersuchung der graduellen Veränderungen vernachlässigt werden; soll dennoch eine solche Untersuchung erfolgen, wird versucht, potentiell neue Cluster gemäß dem in Abschnitt 5.6 vorgestellten Vorgehen mit bereits vorhandenen Clustern zu vereinigen.

Handelt es sich um ein eigenständiges Cluster, muss evaluiert werden, ob durch dieses potentiell neue Cluster $C_{i_{pot}}$ eine ausreichende Zahl an Objekten absorbiert wird; dies wird mit Hilfe des zuvor definierten Parameters α^{A} ermittelt. Gilt die der Bedingung aus (5.14) angelehnte Forderung

$$n_{i_{pot}}^{\alpha^{\mathrm{A}}} \geq \beta_{nC}^{\min} n_{\min}^{\alpha^{\mathrm{A}}}, \tag{5.18}$$

so besteht die Möglichkeit, dass an dieser Stelle ein neues Cluster entsteht bzw. entstanden ist. Wird die Forderung nicht erfüllt, ist dieses Cluster zu verwerfen. Da bei der Wahl von $\beta_{nC}^{\min}$ sichergestellt werden muss, dass auch zukünftig entstehende Cluster Beachtung finden, wird ein zweiter Parameter $\beta_{nC} \in [\beta_{nC}^{\min}, 1]$ benötigt, über den die Mindestgröße eines Clusters definiert wird, das als bereits vorhanden, d.h. nicht mehr in der Entstehung befindlich, angenommen werden kann:

$$n_{i_{pot}}^{\alpha^{\mathrm{A}}} \geq \beta_{nC} n_{\min}^{\alpha^{\mathrm{A}}}. \tag{5.19}$$

Wird die Forderung in (5.18) erfüllt, nicht jedoch die Bedingung für ein entstandenes Cluster in (5.19), so kann es sich bei dem ermittelten Cluster lediglich um ein potentiell entstehendes Cluster handeln. Sind hingegen beide Bedigungen erfüllt, kann das Cluster bzgl. seiner Anzahl an absorbierten Objekten als bereits entstanden betrachtet werden. Bei der Wahl von $\beta_{nC}^{\min}$ und β_{nC} bleibt zu beachten, dass Trends möglichst früh aufgedeckt werden sollen, somit, wie zuvor beschrieben, ein geringer $\beta_{nC}^{\min}$-Wert empfehlenswert ist. Gerade im Marketing-Kontext erweist es sich jedoch als entscheidend, ein Cluster nicht zu früh als eigenständiges Segment zu betrachten, da erst ab einer gewissen Größe die Bearbeitung eines Segments wirtschaftlich sinnvoll erscheint; als Konsequenz empfiehlt sich die Annahme eines vergleichsweise hohen Wert für β_{nC}.

Die Anzahl der durch ein Cluster absorbierten Objekte reicht als einziges Kriterium jedoch nicht aus, um die Entstehung eines neuen Clusters zu sichern; vielmehr muss eine gewisse Dichte innerhalb der Objekte gefordert werden. Auf diese Weise wird gewährleistet, dass es sich nicht um beliebig zusammengefasste Objekte handelt. Dies geschieht auf Basis der von Gath und Geva (1989) eingeführten Partitionsdichte als lokales Gütemaß (vgl. (3.16)). Da im Falle runder Cluster sowie der Erweiterungen der Analyse bei ellipsoiden Clustern die Kovarianzmatrizen nicht explizit bekannt sind (vgl. Kapitel 4), kann die Partitionsdichte je Cluster bei der

possibilistischen Analyse aufgrund der in (3.11) eingeführten Beziehung $\eta_i = \sqrt[p]{\det(\Sigma_i)}$ dabei näherungsweise bestimmt werden als

$$\hat{PD}_i = \frac{\text{card}\left([U_i]_{0.5}\right)}{\sqrt{\eta_i^p}}. \tag{5.20}$$

Auf diese Weise entfällt die gesonderte Berechnung der entsprechenden Fuzzy-Kovarianzmatrizen. Im Falle einer probabilistischen Analyse stellt diese Annäherung keinerlei Einschränkung dar, da durch den wahrscheinlichkeitstheoretischen Ansatz i.d.R. keine echten Ausreißer möglich sind. Es bleibt jedoch zu beachten, dass es sich bei der Berechnung nach (5.20) nicht um die exakten Werte handeln kann, da die η_i lediglich zu Beginn und zusätzlich maximal einmalig im Verlauf der Analyse bestimmt, die Σ_i bzw. die zugehörigen Normmatrizen A_i hingegen kontinuierlich angepasst werden (vgl. Abschnitt 5.3.2). Für einen Vergleich der Werte zur Identifikation potentiell neuer Cluster reicht diese Annäherung jedoch aus.

Analog zur Überprüfung der Clustergröße erfolgt die Prüfung der Dichte in zwei Etappen. Zunächst wird die Einhaltung einer Mindestdichte

$$PD_{i_{pot}} \geq \beta_{nC}^{\min} PD_{\min} \tag{5.21}$$

für entstehende Cluster geprüft, wobei $PD_{\min} = \min\{PD_i | i \in \{1, \ldots, c\}\}$. Wird diese Mindestdichte erreicht, muss weiterhin überprüft werden, inwiefern es sich um ein bereits entstandenes Cluster handelt:

$$PD_{i_{pot}} \geq \beta_{nC} PD_{\min}. \tag{5.22}$$

Erst nach Überprüfung dieser Bedingung wird eine abschließende Aussage möglich.

Alternativ zur Verwendung der Partitionsdichte kann auch die in Abschnitt 5.3.2 verwendete Fuzzykardinalität verwendet werden. Die Partitionsdichte bietet aber beim Vergleich mit bereits existierenden Clustern den Vorteil, dass die Struktur im Kern des Clusters im Verhältnis zur Clusterausdehnung berücksichtigt und damit ein genauerer Vergleich ermöglicht wird.

Zusammenfassend gilt, dass für ein als potentiell entstehend identifiziertes Cluster drei Fälle unterschieden werden können:

1. *Neues Cluster*: Eine gemäß (5.19) ausreichende Anzahl an Objekten wird durch das Cluster absorbiert. Außerdem weist das Cluster eine nach (5.22) als ausreichend zu erachtende Dichte auf.
2. *Potentiell entstehendes Cluster*: Das Cluster besitzt nicht die Struktur eines bereits entstandenen neuen Clusters, es wird jedoch eine gemäß (5.18) minimale Anzahl an Objekten durch das Cluster absorbiert. Ferner weist das Cluster eine nach (5.21) minimal nötige Dichte auf.
3. *Kein zu beachtendes Cluster*: Erfüllt das Cluster nicht die nach (5.18) geforderte minimale Anzahl an Objekten oder ist die Dichte des Clusters gemäß (5.21) nicht als minimal ausreichend zu erachten, findet das Cluster keine weitere Beachtung und wird verworfen.

Wird ein potentiell entstehendes Cluster aufgedeckt, kann es für zukünftige Perioden Relevanz besitzen. In diesem Fall werden in den Folgeperioden die Veränderungen bzgl. Größe und Dichte zum jeweiligen Zeitpunkt benötigt (vgl. Definition 5.2). Die Analyse der Entwicklung dieser

Faktoren im Zweitverlauf muss solange erfolgen, bis das Cluster entweder als neues Cluster anerkannt oder aber verworfen wird. In den folgenden erneuten Analysen bzgl. der Gesamtstruktur sind diese Cluster als eigenständige Cluster zu behandeln, damit die Objektstruktur korrekt erkannt wird sowie vorhandene Cluster nicht durch die resultierenden Häufungen von Ausreißern beeinflusst werden. Die potentiell entstehenden Cluster müssen dabei jedoch als solche markiert werden.

Wird ein bereits existierendes neues Cluster erkannt, so ist dieses ab sofort als vollwertiges Cluster zu betrachten, das in den Folgeperioden analog zu den übrigen vorhandenen Clustern bzgl. seiner graduellen und abrupten Änderungen genauer untersucht werden muss.

5.4.2. Untersuchung der Entwicklung von Ausreißerclustern

Liegen aus vorangegangen Perioden Ausreißercluster vor, die bisher nicht die in Abschnitt 5.4.1 geannten Kriterien bzgl. Größe und Dichte erfüllen, um als eigenes Cluster anerkannt zu werden, muss die Entwicklung dieser Cluster gesondert betrachtet werden. Die Analyse der graduellen Änderungen der Ausreißercluster erfolgt jeweils vor der Analyse weiterer abrupter Veränderungen im Rahmen der allgemeinen Analyse gradueller Unterschiede in der Clusterstruktur. Der Schwerpunkt hierbei liegt auf der Betrachtung der Änderungen bzgl. Objektzahl und Clusterdichte der Ausreißercluster, um ihr Wachstum zu verdeutlichen (vgl. Definition 5.2). Auch die Analyse weiterer lokaler Maße zur Untersuchung gradueller Unterschiede nach Abschnitt 5.3 zur besseren Nachvollziehbarkeit der Veränderungen ist sinnvoll.

Entsprechend erfolgt zunächst die Überprüfung der aktuell durch das Ausreißercluster absorbierten Objekte. Anhand dieser kann gemäß den in Abschnitt 5.4.1 angegebenen Kriterien überprüft werden, ob Clustergröße und -dichte bereits ausreichen, um ein Cluster als eigenständig anzusehen. Ist dies nicht der Fall, muss analysiert werden, inwiefern das Cluster zu verwerfen oder aber ab wann eine ausreichende Erfüllung der genannten Kriterien zu erwarten ist. Unterschreitet ein Cluster eine der Grenzen bzgl. Mindestgröße und -dichte, um als potentiell entstehendes Cluster zu gelten, ist es zu verwerfen. Gleiches gilt, wenn bereits zum ersten Analysezeitpunkt nach Aufdecken des potentiellen Clusters (nahezu) keine Objekte mehr durch das Cluster absorbiert werden. Mögliche Ursachen für das kurzzeitige Auftreten einzelner Cluster können temporäre Anlässe sein: Im Marketingkontext kann das Kundenverhalten z.B. kurzfristig durch gezielte Marketing-Aktionen wie Sonderangebote oder Prämienangebote beeinflusst worden sein.

Ein Cluster wird außerdem verworfen, wenn seine Größe oder Dichte über einen festgelegten Zeitraum $t_{nC}^{\max}$ kein ausreichendes Wachstum oder sogar eine rückläufige Entwicklung aufweist, d.h., das Cluster verschwindet vor seiner vollständigen Ausformung wieder. Anhand der in den einzelnen Perioden ermittelten Parameter $\Delta n_{t_i}^{\alpha^{\mathrm{A}}}$ und ΔPD_{t_i} kann die Entwicklung bzgl. der Clustergröße und -dichte nachvollzogen werden. Wird beim Wachstum ein vorgegebener Grenzwert $\delta_{\min}$ für die vorgegebene Zeitdauer $t_{nC}^{\max}$ unterschritten, so wird das Cluster verworfen, d.h., wenn gilt

$$\frac{\Delta n_{(t-t_{nC}^{\max}+\Delta t)_i}^{\alpha^{\mathrm{A}}}}{\Delta n_{(t-t_{nC}^{\max})_i}^{\alpha^{\mathrm{A}}}} < \delta_{\min} \wedge \ldots \wedge \frac{\Delta n_{t_i}^{\alpha^{\mathrm{A}}}}{\Delta n_{(t-\Delta t)_i}^{\alpha^{\mathrm{A}}}} < \delta_{\min}$$

bzw.

$$\frac{\Delta PD_{(t-t_{nC}^{\max}+\Delta t)_i}}{\Delta PD_{(t-t_{nC}^{\max})_i}} < \delta_{\min} \wedge \ldots \wedge \frac{\Delta PD_{t_i}}{\Delta PD_{(t-\Delta t)_i}} < \delta_{\min}.$$

Wächst das Ausreißercluster, d.h., ist über mehrere Perioden eine Zunahme der Größe oder der Dichte zu verzeichnen, ohne dass es als ausreichend eigenständig anzusehen ist, sollte im Sinne des prädiktiven Ziels der Analyse die Voraussage über die bei der aktuellen Entwicklung benötigte Zeitdauer bis zum Erreichen eines vollwertigen Clusters vorgenommen werden. Die Prognose erfolgt unter Anwendung eines geeigneten Regressionsmodells; die Wahl des Regressionsmodells ist dabei abhängig von der Verteilung der Vergangenheitsdaten und vom Anwender festzulegen.

5.4.3. Veranschaulichung des Vorgehens anhand von künstlichen Daten

Das Vorgehen zur Analyse neu entstehender Cluster soll im Folgenden an einem einfachen Beispiel veranschaulicht werden. Analog zu den vorigen Beispielen wurde ein zweidimensionaler Datensatz mit elliptischen Clustern normalverteilter Daten kreiert. Die Basis stellten drei vorhandene Cluster mit 100 neuen Objekten je Periode dar, insgesamt wurden sieben relevante Perioden betrachtet. Die Zeitfensterlänge betrug wie in den vorangegangenen Beispielen $\tau = 3$, die Analysefrequenz wurde auf $\Delta t = 1$ festgestetzt. Auf diese Weise waren insgesamt fünf Analysen möglich, in der jedes der vorhandenen Cluster aus jeweils 300 Objekten bestand. Zu den drei präsenten Clustern wurde ein viertes, in der Entstehung befindliches Cluster eingeführt. In der ersten Periode lagen noch keinerlei Daten vor, in den beiden Folgeperioden wuchs die Zahl je Periode zunächst langsam um jeweils zehn Objekte an. In den letzten vier Perioden wurde dann ein kontinuierliches, lineares Wachstum um jeweils 20 Objekte je Periode unterstellt, so dass in der siebten Phase insgesamt 100 neue Objekte hinzugefügt wurden (vgl. Tabelle 5.11). Die Ausgangswerte für die Clusterzentren und die Kovarianzmatrizen der einzelnen Cluster bei der Datengenerierung sind in Tabelle 5.12 gegeben.

Cluster	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$
Cluster 1	100	100	100	100	100	100	100
Cluster 2	100	100	100	100	100	100	100
Cluster 3	100	100	100	100	100	100	100
Cluster 4	0	10	20	40	60	80	100

Tabelle 5.11.: Neue Objekte je Periode und Cluster

Für die Analyse wurde der Gustafson-Kessel-Algorithmus für ellipsoide Cluster unter Einbeziehung der Dreiecksbeziehung der Distanzen verwendet (vgl. Abschnitt 4.3.3, Algorithmus 4.7), wobei für den α-Schnitt $\alpha = 0.5$ und als Absorbierungsgrenzwert $\alpha^{\mathrm{A}} = 0.125$ gewählt wurden. Außerdem wurden die η_i während jeder Analyse einmalig neu berechnet (vgl. Abschnitt 3.4), um den Einfluss freier, d.h. nicht absorbierter Objekte auf die übrigen Cluster zu reduzieren. Als Minimalparameter für das Erkennen sich in der Entstehung befindender Cluster wurde

Cluster	Zentrum	Kovarianzmatrix
Cluster 1	$\begin{pmatrix} 2 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$
Cluster 2	$\begin{pmatrix} 11 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 1.5 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$
Cluster 3	$\begin{pmatrix} 12 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 0.8 & -0.5 \\ -0.5 & 0.6 \end{pmatrix}$
Cluster 4	$\begin{pmatrix} 1 \\ 13 \end{pmatrix}$	$\begin{pmatrix} 1.0 & -0.2 \\ -0.2 & 0.4 \end{pmatrix}$

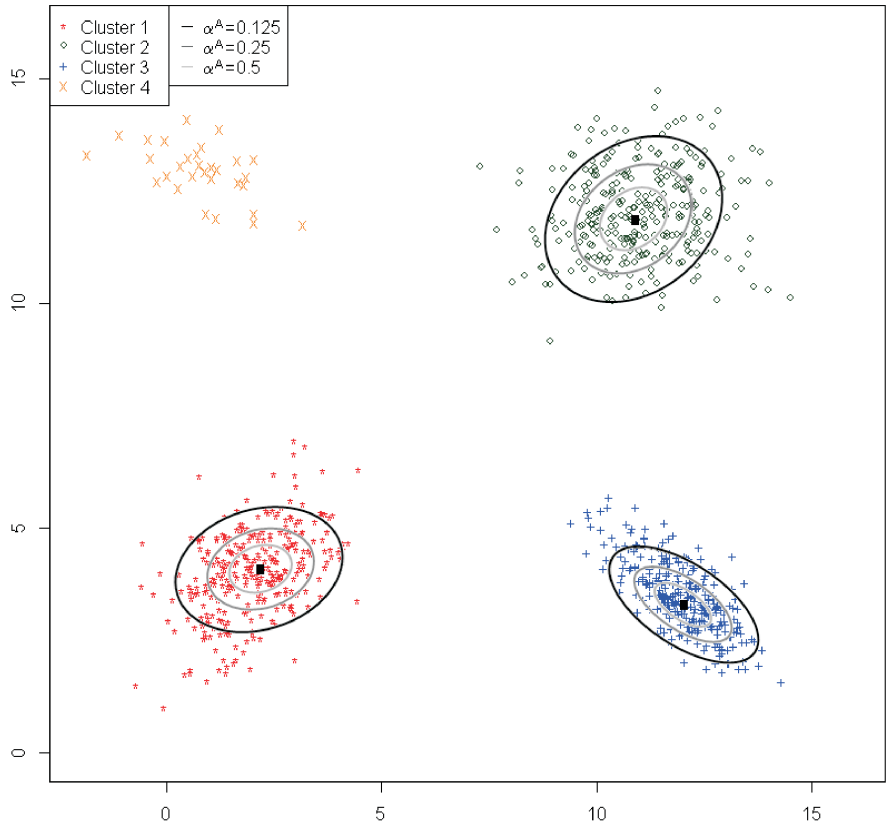
Tabelle 5.12.: Vorgegebene Parameter

$\beta_{nC}^{\min} = 0.3$ festgesetzt, für ausreichend große und dichte neue Cluster $\beta_{nC} = 0.7$. Aufgrund der relativ hohen Wahl für $\beta_{nC}^{\min}$ werden wachsende Cluster nicht allzu früh erkannt, wodurch die Problematik einer zu hohen Wahl von $\beta_{nC}^{\min}$ betont werden soll.

In Abbildung 5.8 sind die Ergebnisse der einzelnen Zeitpunkte grafisch dargestellt. Zum Zeitpunkt $t = 3$ (Abbildung 5.8a) werden die erwarteten drei Cluster wiedergegeben. Die äußeren Ellipsen repräsentieren dabei die harten Kovarianzmatrizen für $\alpha^{\mathrm{A}} = 0.125$, die inneren Ellipsen dienen der besseren Veranschaulichung basierend auf denselben Matrizen bei Zugehörigkeitsgraden von 0.25 bzw. 0.5. Insgesamt werden zu diesem Zeitpunkt 244 Ausreißer identifiziert. Dem neu entstehenden Cluster sind dabei 30 Ausreißer zuzuordnen, die übrigen 214 sind um die vorhandenen Cluster verteilt. Nach (5.14) reicht die Ausreißerzahl zwar zur möglichen Entstehung eines neuen Clusters aus ($214 > 0.3 \cdot 218$, wobei 218 die Anzahl der durch das erste Cluster absorbierten Objekte angibt), es wird jedoch kein neues Cluster aufgedeckt, da sich die Mehrheit der Objekte im direkten Umfeld bereits vorhandener Cluster befindet.

Im darauffolgenden Zeitfenster zum Zeitpunkt $t = 4$, dargestellt in Abbildung 5.8b, ist ein Wachstum im vierten Cluster festzustellen, während die übrigen Cluster nahezu unverändert bleiben. Dieser Befund wird auch durch die lokalen Maße zur Clusterstruktur gemäß Abschnitt 5.3 bestätigt, in denen nahezu keine signifikanten Änderungen nachzuweisen sind. In diesem Zeitfenster werden aufgrund des zunehmenden vierten Clusters bereits 288 Ausreißer identifiziert, von denen sich 218 wiederum im direkten Umfeld der vorhandenen Cluster befinden, 70 hingegen zu dem neu entstehenden Cluster gehören. Die Anzahl der durch das mögliche Ausreißercluster absorbierten Objekte ist weiterhin nicht ausreichend, um das Cluster nach (5.14) als potentiell entstehendes Cluster aufzudecken. Dies liegt insbesondere an dem relativ hohen Wert von $\beta_{nC}^{\min} = 0.3$, so dass $70 < 0.3 \cdot 211$ gilt; 211 ist die minimale Anzahl an Objekten, die durch eines der vorhandenen Cluster absorbiert wird. Durch die Wahl eines geringeren Wertes für den Parameter $\beta_{nC}^{\min}$ würde das entstehende Cluster bereits erkannt.

Zum Zeitpunkt $t = 5$ (Abbildung 5.8c) wird das vierte Cluster erstmals als mögliches entstehendes Cluster erkannt; im aktuellen Zeitfenster besteht das Cluster aus 120 Objekten für die Perioden 3-5, insgesamt werden ebenfalls 288 Ausreißer identifiziert. Durch den Vergleich der verschiedenen möglichen Werte auf Basis der Parameter $\beta_{nC}^{\min}$ und β_{nC} mit den jeweils minimalen vorhandenen Werte folgt, dass es sich um ein potentiell entstehendes Cluster handelt; die entsprechenden Werte sind in Tabelle 5.13 wiedergegeben. Die Fuzzy-Kardinalitäten sind zur besseren Vergleichbarkeit ebenfalls aufgeführt.



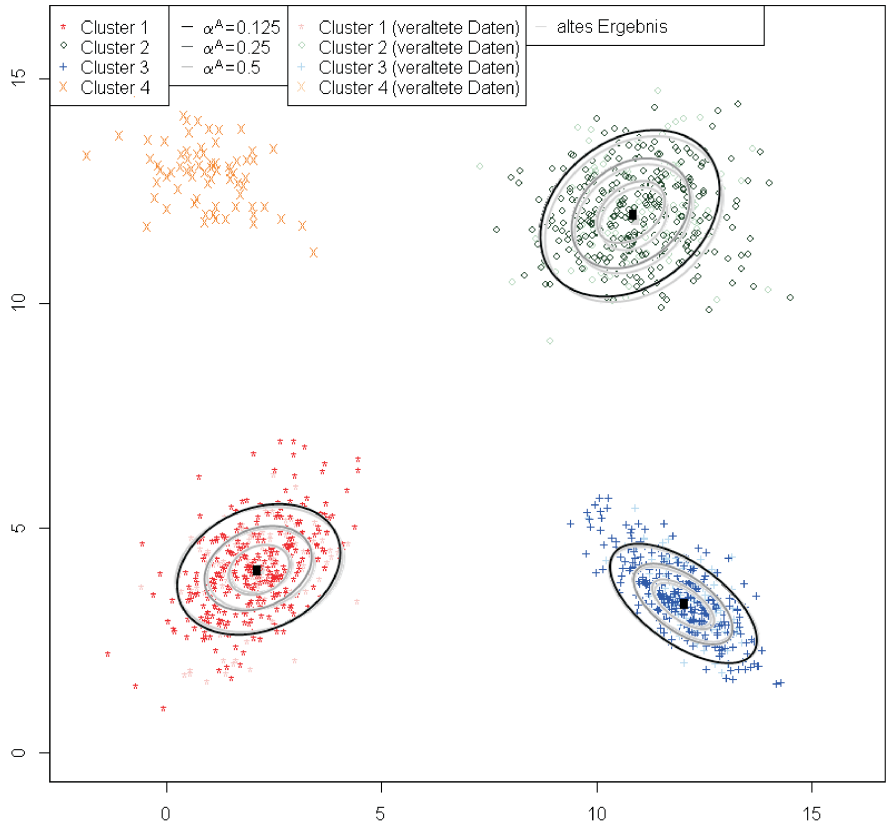
(a) Clusterstruktur für den Zeitpunkt $t = 3$

Abbildung 5.8.: Beispiel für Neubildung eines Clusters

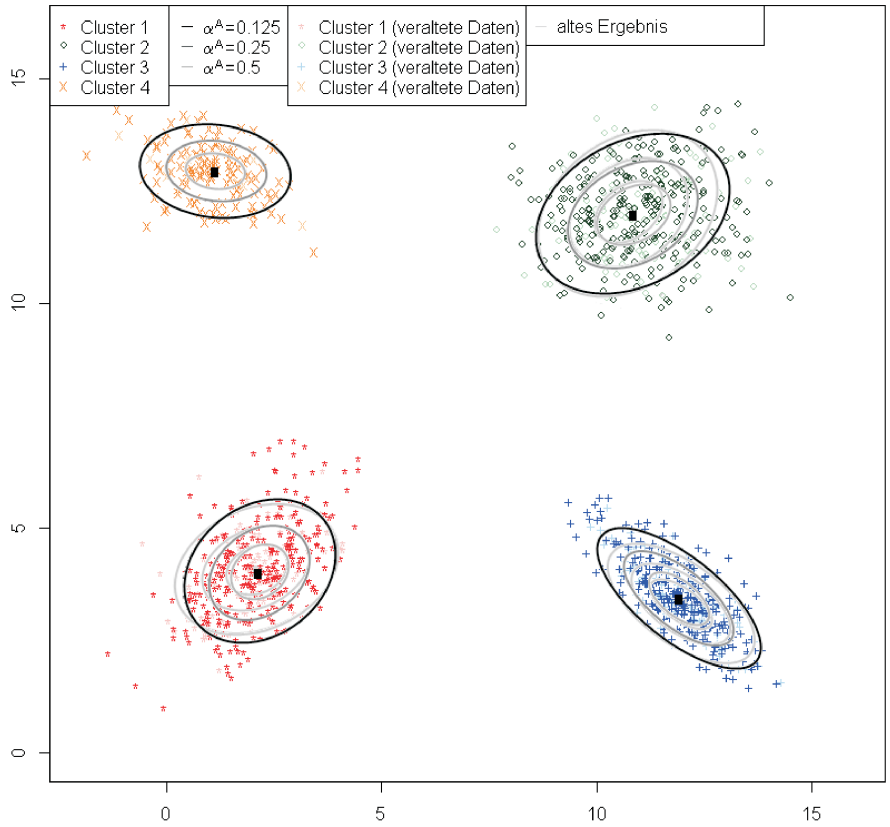
	n_4	PD_4	card $([U_4]_{0.125})$
Cluster 4	90	34.959	37.960
Min. Werte	240	72.307	86.083

Tabelle 5.13.: Vergleichswerte für neues Cluster zum Zeitpunkt $t = 5$

Das Aufdecken des möglichen vierten Clusters und die entsprechende Berücksichtigung bei der Clusteranalyse zum Zeitpunkt $t = 5$ führt zur Repräsentation aller vier Cluster im anschließenden Zeitfenster (vgl. Abbildung 5.8d). Das neue Cluster 4 weist zum Zeitpunkt $t = 6$ weiterhin ein Wachstum auf, angezeigt durch zunehmende Kardinalitäten. Es absorbiert jedoch immer noch eine verhältnismäßig geringe Anzahl an Objekten, so dass es weiterhin als wachsendes Cluster identifiziert wird (vgl. Tabelle 5.14). Die übrigen Cluster sind nahezu unverändert. Wie in Tabelle 5.14 deutlich wird, ist die Partitionsdichte des neuen Clusters bereits mehr als ausreichend, um als eigenständiges Cluster angesehen zu werden; lediglich die Anzahl absorbierter Objekte liegt noch knapp unter dem derzeitigen Grenzwert von $0.7 \cdot 204 = 142.8$. Auf Basis der wenigen vorhandenen Werte der Zeitfenster $t = 5$ und $t = 6$ kann nun unter Annahme linearer Zuwachsraten bestimmt werden, wann eine ausreichende Objektzahl innerhalb des

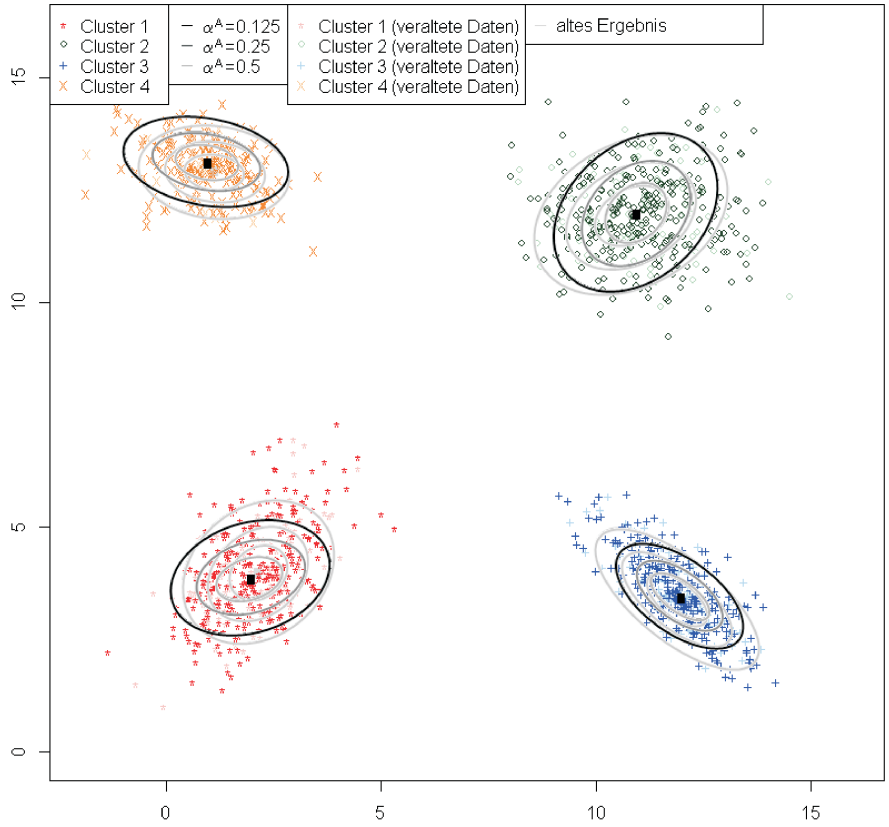


(b) Clusterstruktur für den Zeitpunkt $t = 4$

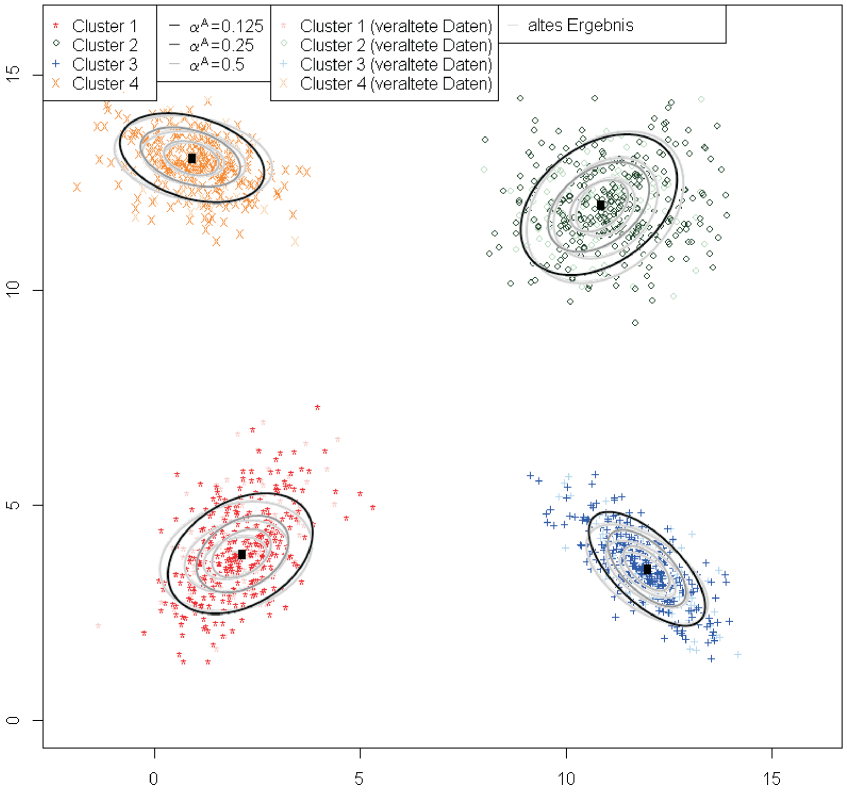


(c) Clusterstruktur für den Zeitpunkt $t = 5$

Abbildung 5.8.: Beispiel für Neubildung eines Clusters



(d) Clusterstruktur für den Zeitpunkt $t = 6$



(e) Clusterstruktur für den Zeitpunkt $t = 7$

Abbildung 5.8.: Beispiel für Neubildung eines Clusters

	n_4	PD_4	$\text{card}\left([U_4]_{0.125}\right)$
Cluster 4	134	105.708	54.323
Min. Werte	204	78.843	73.820

Tabelle 5.14.: Vergleichswerte für neues Cluster zum Zeitpunkt $t = 6$

Clusters zu erwarten ist[21]. Entsprechend lässt sich anhand der gegebenen Werte die Gleichung

$$n(t) = 44t - 130$$

aufstellen, so dass ab dem Zeitpunkt $t = 7$ mit einer Objektzahl von $n(7) = 178$ ein ausreichend großes Cluster zu erwarten ist. Betrachtet man das letzte Zeitfenster in Abbildung 5.8e, so wird das Cluster als eigenständiges neues Cluster erkannt; die Anzahl absorbierter Objekte liegt aufgrund der sehr einfachen Schätzfunktion für $n(t)$ leicht über dem erwarteten Wert, wie Tabelle 5.15 zu entnehmen ist. Die übrigen Cluster weisen keine signifikanten Veränderungen gegenüber der vorigen Periode auf.

	n_4	PD_4	$\text{card}\left([U_4]_{0.125}\right)$
Cluster 4	187	111.629	70.115
Min. Werte	208	91.923	76.835

Tabelle 5.15.: Vergleichswerte für neues Cluster zum Zeitpunkt $t = 7$

Anhand des Beispiels konnte verdeutlich werden, wie die Entstehung eines neuen Clusters im dynamischen Kontext nachvollzogen werden kann. Dabei ist die geeignete Wahl der Parameter $\beta_{nC}^{\min}$ für ein potentiell entstehendes und β_{nC} für ein bereits entstandenes neues Clusters elementar, um negative Auswirkungen im Rahmen der Analyse möglichst gering zu halten.

5.5. Eliminierung veralteter Cluster

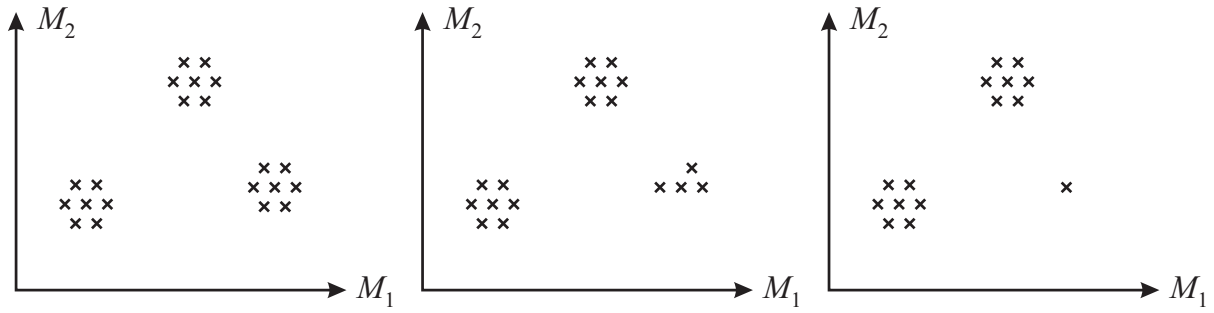
Durch Veränderungen von Objekteigenschaften im Laufe der Zeit kann es zum Sterben einzelner Cluster kommen; diese Cluster müssen im Zeitablauf eliminiert werden, da sie über keinerlei Bedeutung für die Erklärung einer Clusterstruktur mehr verfügen. Im ökonomischen Kontext ist das Erkennen der gefährdeten Cluster elementar, da nur durch das frühzeitige Aufdecken der abnehmenden Cluster einer negativen Entwicklung entgegengewirkt werden kann, abhängig davon, ob eine Aufrechterhaltung eines Segments erfolgsversprechend erscheint, oder andernfalls Vorbereitungen für die Elimination dieses Segments getroffen werden sollten.

Im Folgenden werden zunächst die Parameter aufgeführt, an denen eine solche Entwicklung nachvollzogen werden kann, bevor das generelle Vorgehen im Anschluss anhand eines künstlichen Beispiels veranschaulicht wird.

[21] Dieses Vorgehen ist aufgrund der Einfachheit des Beispiels in diesem Fall stark simplifiziert, da nur zwei Zeitpunkte die Grundlage bilden.

5.5.1. Vorgehen zum Erkennen zu eliminierender Cluster

Wie bei der Neuentstehung eines Clusters erfolgt auch sein Sterben generell nicht ohne Vorwarnung. Zur graphischen Darstellung kann das Beispiel in Abbildung 5.7 hinzugezogen werden, jedoch in zeitlich umgekehrter Reihenfolge (vgl. Abbildung 5.9). Zum Zeitpunkt t_1 (Abbildung 5.9a) lassen sich drei vollwertige Cluster erkennen. Eine Periode später kann bei einem Cluster bereits ein Rückgang bzgl. Größe und Dichte verzeichnet werden (Abbildung 5.9b); inwiefern dieses Cluster noch die Eigenschaften eines ausreichend selbstständigen Clusters erfüllt, bedarf einer separaten Überprüfung. Die graduellen Veränderungen der Größe und der Dichte implizieren jedoch in jedem Fall bereits eine rückläufige Entwicklung dieses Clusters. Zum Zeitpunkt t_3 (Abbildung 5.9c) ist das Cluster nahezu vollständig abhanden gekommen und muss aus der bekannten Clusterstruktur eliminiert werden.



(a) Zeitpunkt t_1 (b) Zeitpunkt t_2 (c) Zeitpunkt t_3

Abbildung 5.9.: Sterben eines Clusters

Um in höherdimensionierten Räumen bei weniger eindeutigen Strukturen zu überprüfen, ob Cluster bzgl. ihrer Größe oder Dichte einen negativen Trend vorweisen, werden entsprechende Kennzahlen benötigt, die Aufschluss über die Gefährdung eines Clusters geben. Wie schon bei der Neubildung von Clustern (vgl. Abschnitt 5.4) können drei charakteristische Fälle für jedes Cluster unterschieden werden:

1. *Keine Gefährdung*: Das Cluster ist weiterhin existent.
2. *Gefährdung*: Die Entwicklung des Clusters ist rückläufig; eine zukünftige Elimination ist möglich.
3. *Clusterelimination*: Das Cluster existiert nicht länger, es wird aus der allgemeinen Clusterstruktur eliminiert.

Ob für ein Cluster eine potentielle Gefährdung besteht, lässt sich vorab anhand der graduellen Unterschiede innerhalb des Clusters untersuchen (vgl. Abschnitt 5.3): Nur beim Vorhandensein signifikanter Änderungen, die eine mögliche Clusterelimination implizieren, bedarf es einer weiterführenden Untersuchung. Hierzu sind die absolute sowie die Fuzzy-Kardinalität und im Falle ellipsoider Cluster das Fuzzy-Hypervolumen von grundlegender Bedeutung, da ein Rückgang der Kardinalitäten bzw. eine Zunahme des Volumens die Gefährdung eines Clusters anzeigen können. Gilt ein Cluster nach Analyse der graduellen Unterschiede in jedem Fall als ungefährdet, so ist es weiterhin existent (Fall 1). Ebenso gilt ein Cluster nicht als gefährdet, wenn zwar signifikante Änderungen bzgl. der relevanten Maße vorliegen, die Anzahl der neu bzw. insgesamt

absorbierten Objekte und ihre Dichte jedoch weiterhin ausreichen, das Cluster als eigenständig ansehen zu können. Hierzu ist eine separate Prüfung erforderlich.

Zunächst ist die Anzahl im aktuellen Zeitfenster neu absorbierter Objekte zu untersuchen, d.h. die Anzahl der in den letzten Δt Zeiteinheiten des aktuellen Zeitfensters zum Cluster hinzugekommen Objekte. Diese gilt als ausreichend, sofern

$$n_{i_{neu}}^{\alpha^{\mathrm{A}}} \geq \lambda_{\min}^{neu}, \tag{5.23}$$

wobei $\lambda_{\min}^{neu}$ angibt, wie viele Objekte mindestens hinzukommen müssen, um die Anzahl neu absorbierter Objekte als ausreichend zu betrachten. Der Parameter $\lambda_{\min}^{neu}$ ist entweder kontextabhängig festzusetzen oder im Vergleich zu den übrigen (ungefährdeten) Clustern zu bestimmen, beispielsweise relativ zu der durchschnittlichen Anzahl neu absorbierter Objekte aller nach Prüfung der graduellen Änderungen ungefährdeten Cluster. Wird die Bedingung in (5.23) verletzt, so liegt keine ausreichende Neuzuordnung vor. In diesem Fall treten im aktuellen Zeitfenster nur wenige neue Objekte auf, die die Existenz des Clusters begründen. Damit muss eine Eliminierung des Clusters grundsätzlich in Erwägung gezogen, das Cluster jedoch nicht automatisch eliminert werden. Vielmehr muss eine weitergehende Untersuchung bzgl. Größe und Dichte des Clusters durchgeführt werden, um den Einfluss zufälliger, kurzfristiger Schwankungen zu reduzieren. Auf diese Weise wird zudem sichergestellt, dass die durch das Cluster absorbierten Objekte aufgrund einer verfrühten Clusterelimination die Lage der übrigen Cluster beim Reclustering nicht beeinflussen (vgl. Abschnitt 5.4).

Neben der Anzahl neu absorbierter Objekte erfolgt die Analyse der Gesamtzahl der durch das Cluster im aktuellen Zeitfenster absorbierten Objekte sowie deren Dichte im Vergleich zu den Clustern, die nach der Untersuchung der graduellen Veränderungen als ungefährdet einzustufen sind. Würden alle Cluster in den Vergleich einbezogen, wäre eine Verfälschung des Ergebnisses möglich, da bei mehreren gefährdeten Clustern die Minimalwerte sich ggf. aufheben und die Gefährdungen nicht deutlich würden. Der Vergleich der Gesamtzahl absorbierter Objekte und deren Dichte erfolgt analog zum Vorgehen beim Aufdecken potentieller neuer Cluster auf Basis der Minimalwerte ungefährdeter Cluster, d.h., ein Cluster ist dann als ungefährdet anzusehen, wenn neben der Bedingung aus (5.23)

$$n_i^{\alpha^{\mathrm{A}}} \geq \beta_{eC} n_{\min}^{\alpha^{\mathrm{A}}} \tag{5.24}$$

und

$$PD_i \geq \beta_{eC} PD_{\min} \tag{5.25}$$

gelten, wobei $n_{\min}^{\alpha^{\mathrm{A}}} = \min\left\{n_i^{\alpha^{\mathrm{A}}} | C_i \text{ ist ungefährdet}\right\}$ und $PD_{\min} = \min\{PD_i | C_i \text{ ist ungefährdet}\}$ die Minimalwerte ungefährdeter Cluster angeben und $\beta_{eC} \in [\beta_{eC}^{\min}, 1]$ der Parameter zur Bestimmung der Mindestgröße bzw. -dichte eigenständiger Cluster ist. Der Parameter $\beta_{eC}^{\min} \in [0, 1]$ gibt die Höhe des minimal geforderte Anteils an, um ein Cluster als weiterhin existent, wenn auch gefährdet, anzusehen. Je nach Kontext ist eine einheitliche Wahl der Parameter zur Analyse der Neubildung von Clustern denkbar, d.h. $\beta_{eC}^{\min} = \beta_{nC}^{\min}$ sowie $\beta_{eC} = \beta_{nC}$. I.d.R. sollte jedoch $\beta_{eC}^{\min} > \beta_{nC}^{\min}$ gewählt werden, da ein gefährdetes Cluster im Gegensatz zum potentiell entstehenden Cluster als noch vorhandenes Cluster gilt.

Wird eine der oben genannten Bediungen verletzt, so ist ein Cluster zumindest gefährdet. Ob es bereits zu eliminieren ist, erfordert weitergehende Untersuchungen. Bei einer unzureichenden Neuzuordnung von Objekten gemäß (5.23), d.h., wenn

$$n_{i_{neu}}^{\alpha^{\mathrm{A}}} < \lambda_{\min}^{neu},$$

wird das Cluster als gefährdet markiert, bevor die übrigen Parameter untersucht werden. Die Markierung kann aufgehoben werden, falls in den Folgeperioden wieder alle Bedingungen zur weiteren Existenz des Clusters erfüllt werden.

Reicht die Gesamtzahl der absorbierten Objekte nicht länger aus, d.h.

$$n_i^{\alpha^{\mathrm{A}}} < \beta_{eC} n_{\min}^{\alpha^{\mathrm{A}}},$$

so ist eine mögliche Eliminierung des Clusters zu überprüfen. Das Cluster ist bzgl. seiner Objektzahl lediglich als gefährdet anzusehen, wenn analog zur Neubildung der Cluster und Bedingung (5.18) gilt

$$n_i^{\alpha^{\mathrm{A}}} \geq \beta_{eC}^{\min} n_{\min}^{\alpha^{\mathrm{A}}}. \tag{5.26}$$

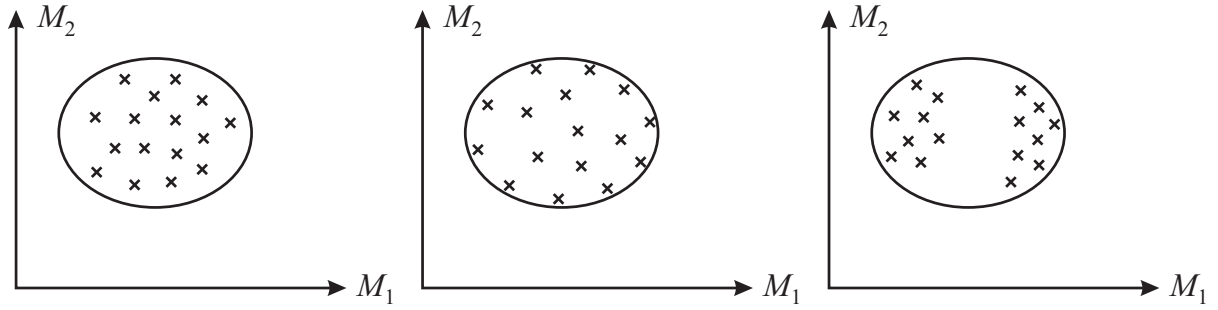
Wird diese Bedingung ebenfalls verletzt, besitzt das Cluster keine ausreichende Relevanz mehr und kann aus der Gesamtstruktur eliminiert werden. Eine solche Elimination ist beispielsweise dann von Bedeutung, wenn ein Kundensegment keine ausreichende Anzahl an Kunden mehr vorweisen kann, um als attraktives Segment weiterhin ökonomisch sinnvoll bearbeitet zu werden.

Ähnliches gilt bei der Evaluierung der Clusterdichte. Die Dichte der absorbierten Objekte reicht nicht länger aus, wenn

$$PD_i < \beta_{eC} PD_{\min} \tag{5.27}$$

gilt. Ist dies der Fall, so muss beachtet werden, dass die Ursache für die Änderung der Clusterdichte nicht zwangsläufig eine mögliche Clusterelimination sein muss: Ebenso kann auch eine Umverteilung der Dichte innerhalb des Clusters und damit eine bevorstehende Clustertrennung den Grund eines Dichterückgangs darstellen. Liegt nach Analyse der graduellen Veränderungen eines Clusters ein negativer Trend bzgl. der Clusterdichte vor, muss entsprechend zunächst überprüft werden, inwiefern eine Clusterteilung anstelle einer -elimination bevorsteht. Abbildung 5.10 verdeutlicht dieses Problem: In Abbildung 5.10a ist ein eigenständiges Cluster gegeben. In den darauffolgenden Abbildungen 5.10b und 5.10c sind zwei Beispiele mit rückläufiger Partitionsdiche des Clusters dargestellt. Während die absolute Kardinalität des Cluster keine signifikante Änderung aufweist, zeigen die Änderungen bzgl. Fuzzy-Kardinalität und Volumen jedoch bereits eine Gefährdung des Clusters an. In Abbildung 5.10b ist eine Gefährdung dieses Clusters deutlich erkennbar, da es sich aufzulösen scheint, wohingegen das Cluster in Abbildung 5.10c vor einer Teilung steht. Entsprechend muss bei einem Rückgang der Fuzzy-Kardinalität und damit einhergehend der Partitionsdichte ohne gleichzeitige signifikante Änderung der absoluten Kardinalität zunächst auf Basis der Änderungen bzgl. der Kompaktheit geprüft werden, inwiefern eine Clustertrennung bevorsteht; das zugehörige Vorgehen wird in Abschnitt 5.7 zur Clustertrennung detailliert erläutert.

Ist keine Teilung des Cluster zu erwarten, erfolgt ein weiterer Vergleich der Partitionsdichte analog zur absoluten Kardinalität mit der minimalen Dichte der nicht gefährdeten Cluster.



(a) Ausgangslage (b) Bevorstehende Elimination (c) Bevorstehende Clustertrennung

Abbildung 5.10.: Implikationen durch Rückgang der Partitionsdichte

Wird die Bedingung

$$PD_i \geq \beta_{eC}^{\min} PD_{\min}$$

verletzt, so ist das Cluster zu eliminieren, ansonsten wird es lediglich als gefährdet markiert.

Zusammenfassend gilt ein Cluster dann als gefährdet, wenn keine ausreichende Neuzuordnung im aktuellen Zeitfenster erfolgt oder die Gesamtzahl der durch das Cluster absorbierten Objekte bzw. ihre Dichte nicht länger ausreicht, um als vollständig eigenständiges Cluster zu gelten.

5.5.2. Untersuchung der Entwicklung gefährdeter Cluster

Bei als gefährdet markierten Clustern ist eine Nachverfolgung im Zeitablauf erforderlich, um frühzeitig Trends aufdecken und mögliche Maßnahmen einleiten zu können. Werden in den Folgeperioden die Bedingungen für ein existierendes Cluster erfüllt, d.h., es werden genügend neue Objekte absorbiert und die Clustergröße und -dichte reichen aus, so kann die Markierung als gefährdetes Cluster aufgehoben werden. Es sollte jedoch zusätzlich kontextabhängig abgeklärt werden, wodurch die Schwankungen verursacht wurden. Handelt es sich beispielsweise um saisonale Schwankungen, kann auch in Zukunft wieder mit einer entsprechenden zeitweisen Modifikation gerechnet werden.

Werden über einen Zeitraum $t_{eC}^{\max}$ nicht ausreichend viele neue Objekte durch ein Cluster absorbiert, d.h.

$$n_{t'_{ineu}}^{\alpha^{\mathrm{A}}} < \lambda_{\min}^{neu} \quad \forall t' \in [t - t_{eC}^{\max}, t]\,, \tag{5.28}$$

so ist das Cluster zu eliminieren, da es keine Bedeutung mehr für für die Clusterstruktur besitzt. Der Parameter $t_{eC}^{\max}$ ist kontextabhängig vom Anwender festzulegen.

Nimmt die Objektzahl oder die Dichte eines Clusters monoton ab, ist analog zur Neubildung von Clustern der voraussichtliche Eliminationszeitpunkt abzuschätzen. Die Prognose erfolgt wiederum unter Anwendung eines geeigneten Regressionsmodells, wobei dieses abhängig von der Verteilung der Vergangenheitsdaten ist und die Wahl des Modells kontextabhängig erfolgt.

Wird ein Cluster eliminiert, so sollten die das Cluster betreffenden Informationen nicht endgültig gelöscht werden (vgl. Crespo und Weber, 2005): Diese Informationen können bei Neubildung eines Clusters an derselben oder ähnlicher Stelle relevant werden. Dieses Verfahren ermöglicht es, ohne zusätzlichen Informationsverlust frühzeitigen Eliminationen entgegenzuwirken. Außerdem können saisonale Schwankungen identifiziert werden, die ein vorübergehendes Unterschreiten der Grenzwerte bzgl. Größe und Dichte eines Clusters beinhalten.

5.5.3. Veranschaulichung des Vorgehens anhand von künstlichen Daten

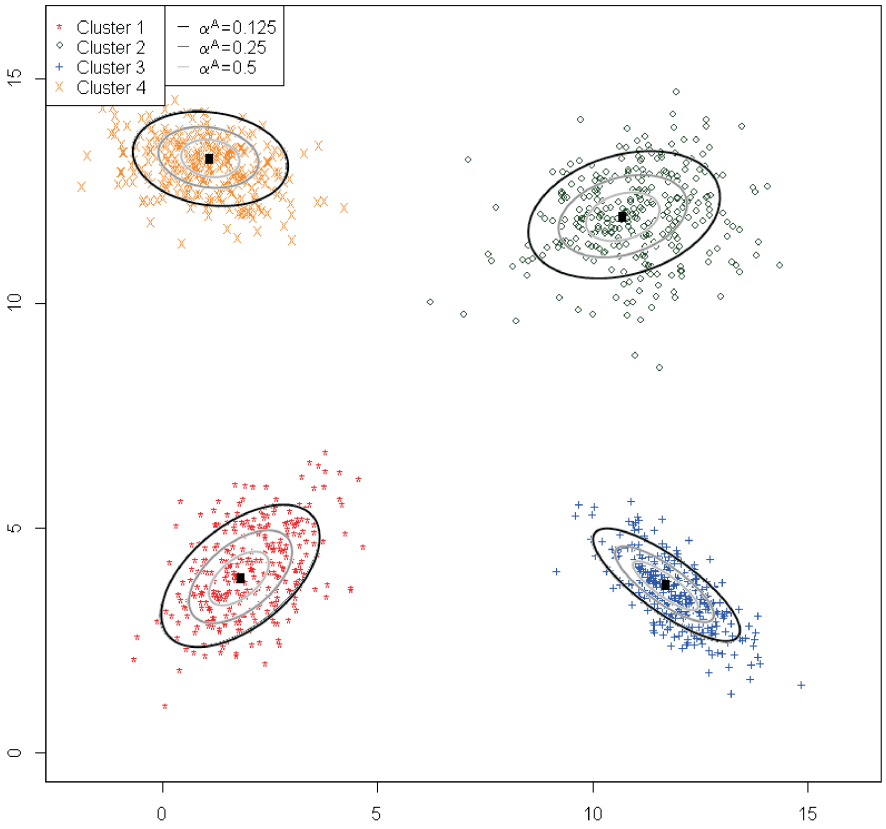
Im Folgenden wird das Vorgehen zum Erkennen gefährdeter Cluster anhand eines vereinfachten Beispiels erläutert. Es wurde ein zweidimensionaler Datensatz normalverteilter Daten kreiert, dessen Basis vier elliptische Cluster darstellten. Zu drei dieser Cluster kamen in jeder Periode konstant 100 Objekte hinzu, d.h., ab der dritten Periode bestand jedes Cluster aus konstant 300 Objekten. Für das vierte Cluster wurde eine rückläufige Entwicklung angenommen: Ab der dritten Periode wurde die Objektzahl, ebenfalls ausgehend von 100 Objekten je Periode, zunächst relativ leicht um 10 Objekte, ab der fünften Periode dann um 20 Objekte je Periode reduziert. Der Beobachtungszeitraum erstreckte sich über insgesamt acht Perioden; dem vierten Cluster konnten somit in der letzten Periode keinerlei neue Objekte mehr zugeordnet werden (vgl. Tabelle 5.16). Die Zeitfensterlänge lag bei $\tau = 3$, als Analysefrequenz wurde entsprechend den vorherigen Beispielen $\Delta t = 1$ gewählt. Die Ausgangswerte für die Clusterzentren und die Kovarianzmatrizen der einzelnen Cluster zur Datengenerierung wurden analog zum Beispiel in Abschnitt 5.4.3, Tabelle 5.12, gewählt.

Cluster	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$
Cluster 1	100	100	100	100	100	100	100	100
Cluster 2	100	100	100	100	100	100	100	100
Cluster 3	100	100	100	100	100	100	100	100
Cluster 4	100	100	90	80	60	40	20	0

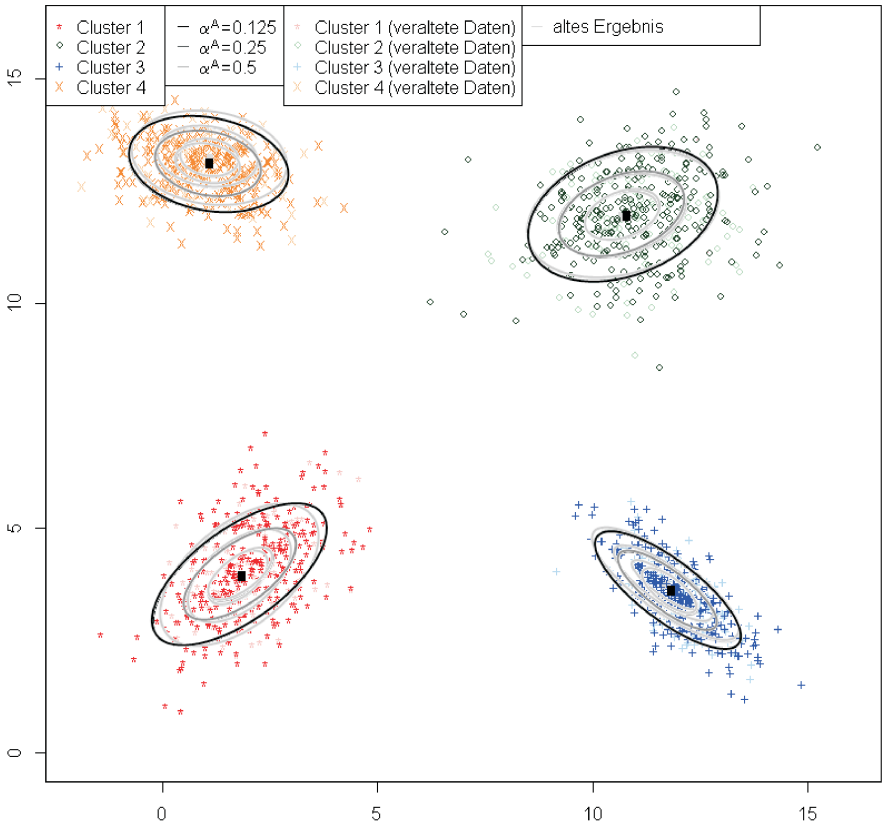
Tabelle 5.16.: Neue Objekte je Periode und Cluster

Die einzelnen Analysen wurden unter Anwendung des Gustafson-Kessel-Algorithmus für ellipsoide Cluster bei Einbeziehung der Dreiecksbeziehung der Distanzen durchgeführt (vgl. Abschnitt 4.3.3, Algorithmus 4.7). Dabei wurde jeweils einmailig eine Neuberechnung der η_i vorgenommen, um den Einfluss freier, nicht absorbierter Objekte, die aufgrund einer Clustereliminination auftreten können, auf die übrigen Cluster zu reduzieren. Für den α-Schnitt wurde analog zu den vorigen Beispielen $\alpha = 0.5$ gewählt, der Grenzwert der Absorbierung wurde auf $\alpha^{\mathrm{A}} = 0.125$ bestimmt. Die Minimalzahl $\lambda_{\min}^{neu}$ wurde auf 60% der durchschnittlichen Anzahl an Objekten festgesetzt, die durch ungefährdete Cluster absorbiert werden. Diese Anzahl durfte in maximal zwei aufeinanderfolgenden Perioden unterschritten werden, ohne dass das Cluster eliminert wurde ($t_{eC}^{\max} = 2$). Für die Bestimmung des Gefährdungsgrades eines Clusters wurden $\beta_{eC} = 0.7$ und $\beta_{eC}^{\min} = 0.5$ gewählt.

Abbildung 5.11 dokumentiert die Analyseergebnisse. Zum Zeitpunkt der ersten Analyse

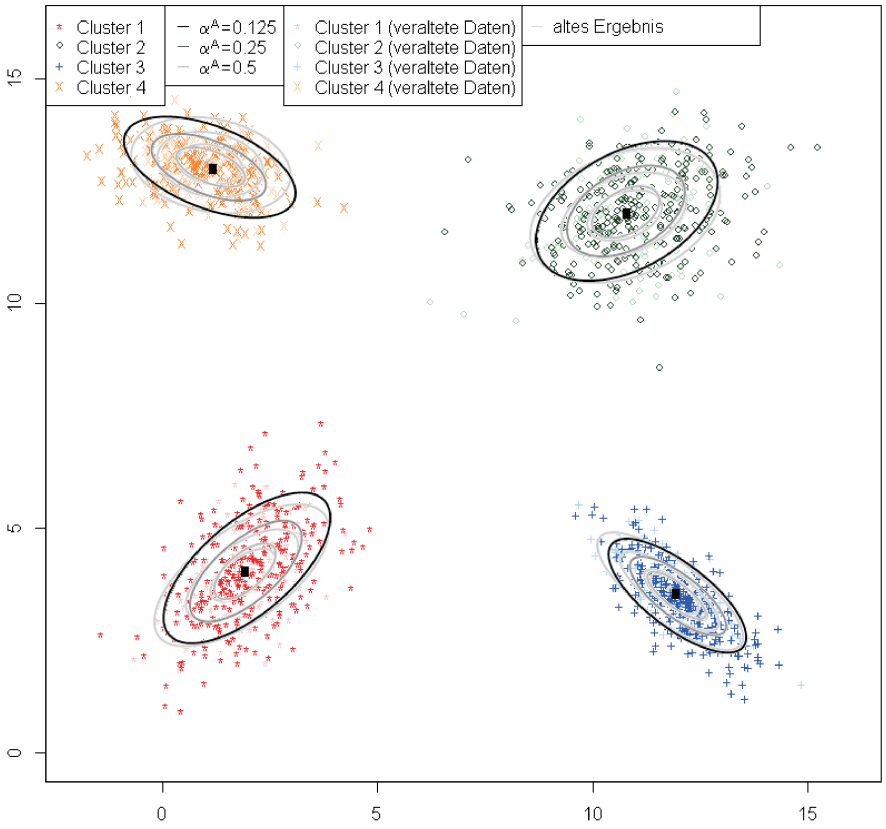


(a) Clusterstruktur für den Zeitpunkt $t = 3$

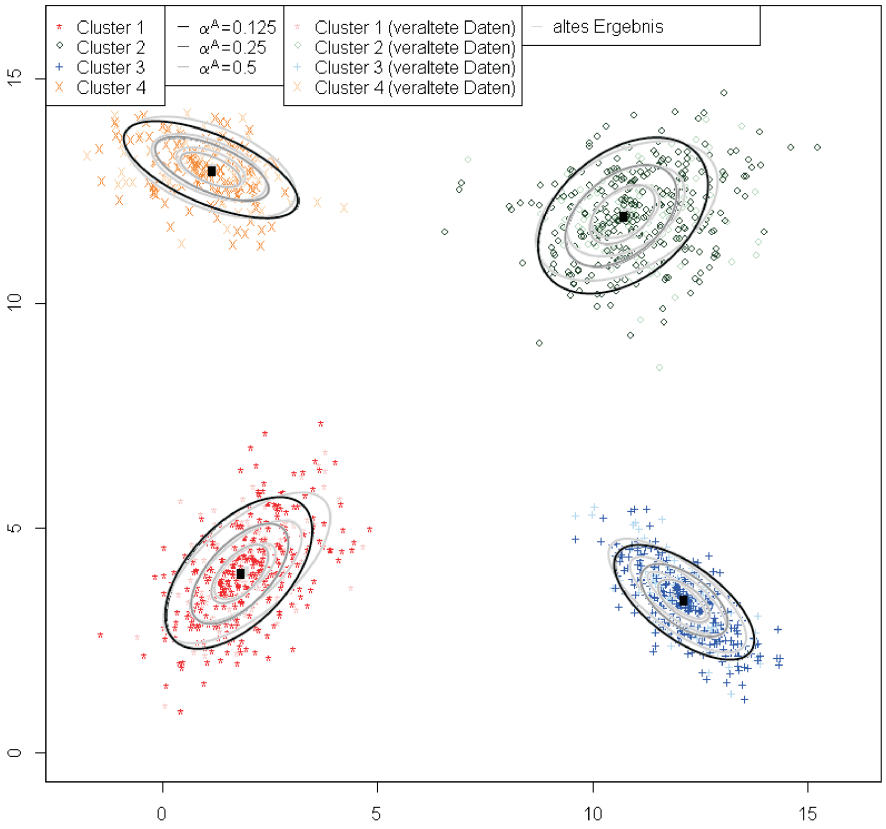


(b) Clusterstruktur für den Zeitpunkt $t = 4$

Abbildung 5.11.: Beispiel für Clusterelimination

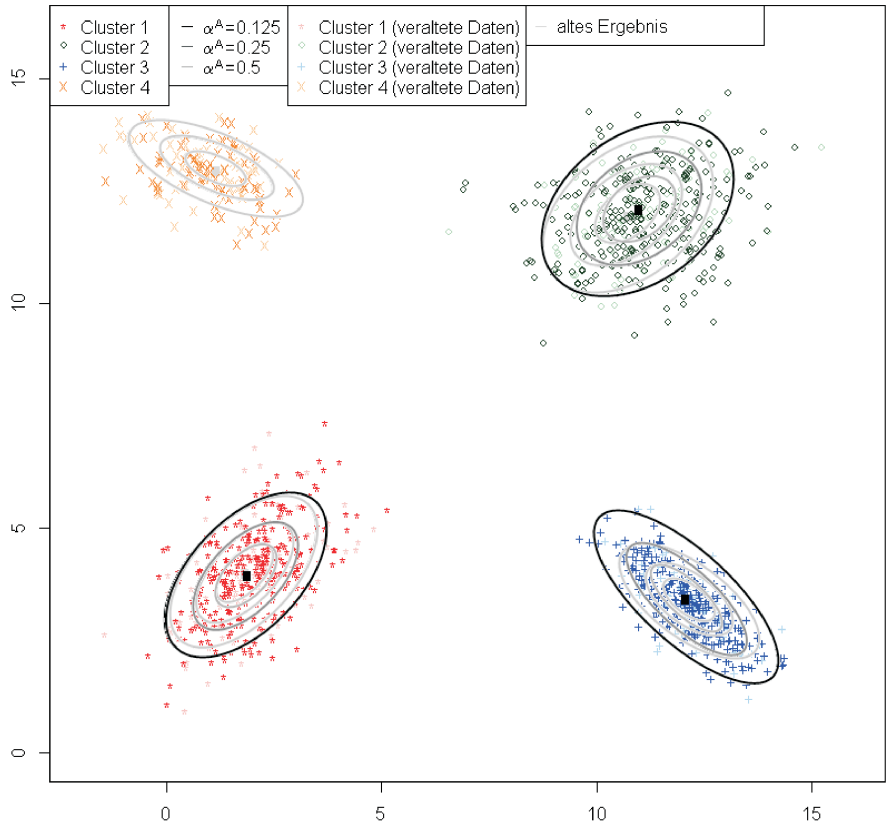


(c) Clusterstruktur für den Zeitpunkt $t = 5$

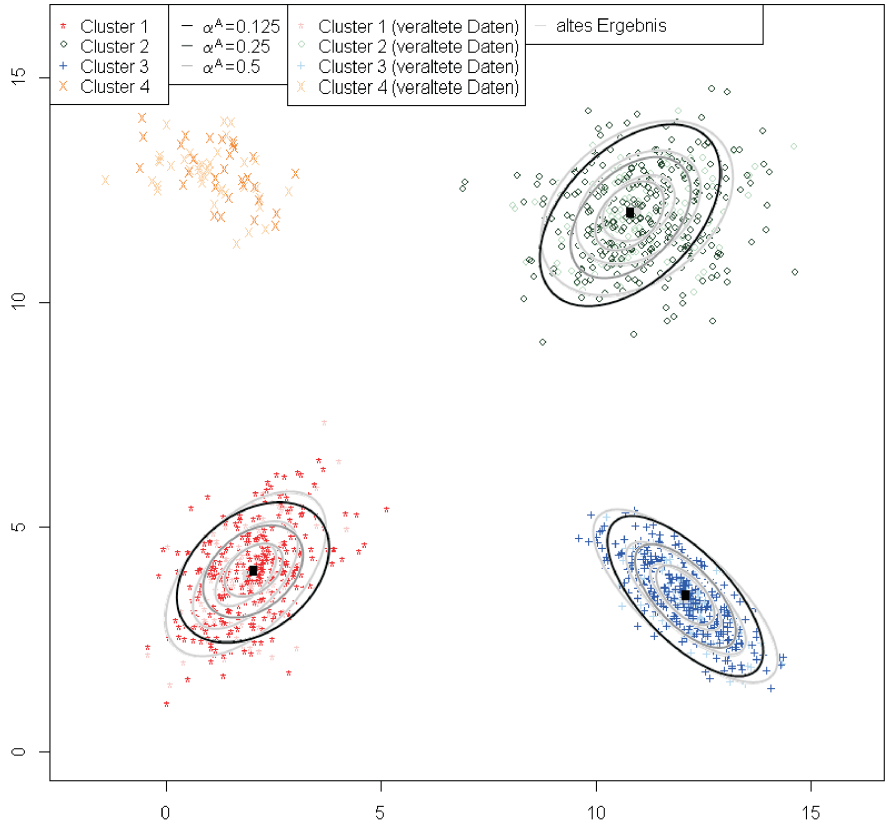


(d) Clusterstruktur für den Zeitpunkt $t = 6$

Abbildung 5.11.: Beispiel für Clusterelimination



(e) Clusterstruktur für den Zeitpunkt $t = 7$



(f) Clusterstruktur für den Zeitpunkt $t = 8$

Abbildung 5.11.: Beispiel für Clusterelimination

($t = 3$, Abbildung 5.11a) werden durch die Fuzzy-Clusteranalyse wie erwartet vier Cluster aufgedeckt. Die etwas geringere Objektzahl des vierten Clusters besitzt für die allgemeine Analyse keine Relevanz. Zum darauffolgenden Zeitpunkt $t = 4$ können bereits Auswirkungen der verringerten Objektzahl auf die Kardinalitäten des vierten Clusters beobachtet werden: Durch die Objektzahlreduktion nehmen die Kardinalitäten deutlich ab, während die übrigen lokalen Parameter zur Messung gradueller Unterschiede gemäß Abschnitt 5.3 keine signifikanten Änderungen aufweisen. Die geringeren Kardinalitäten implizieren bereits die rückläufige Entwicklung; aufgrund der verbliebenen Größe und Dichte des Clusters gilt es jedoch noch nicht als gefährdet. Die lokalen Maße der anderen Cluster zeigen keine signifikanten Änderungen, sie sind nahezu unverändert (vgl. Abbildung 5.11b).

Der Rückgang der Kardinalitäten des vierten Clusters aufgrund der zunehmenden Objektzahlreduktion zeigt sich zum Zeitpunkt $t = 5$ noch deutlicher. Die Entwicklung erzeugt wegen des Fortfalls einzelner Objekte in der clustereigenen Struktur außerdem Schwankungen bzgl. des lokalen Partitionskoeffizienten LPC sowie des Fuzzy-Hypervolumens FHV. Es handelt sich jedoch um zufällige Schwankungen, die außer auf den beliebigen Wegfall einzelner Objekte auf keine konkrete Entwicklung zurückgeführt werden können. Während die übrigen Cluster in ihrer Struktur nahezu unverändert bleiben (vgl. Abbildung 5.11c), gilt Cluster 4 nun als gefährdet, da die Anzahl durch das Cluster neu absorbierter Objekte, $n_{4_{neu}}^{\alpha_A} = 30$, verglichen mit den verbliebenen Clustern und dem Grenzwert $\lambda_{\min}^{neu} = 41.4$ nicht länger ausreicht. Die Gesamtzahl der absorbierten Objekte und die Dichte des Clusters deuten noch nicht auf eine anstehende Clusterlimination hin (vgl. Tabelle 5.17), unter anderem aufgrund der Wahl von $\beta_{eC} = 0.7$; eine höhere Wahl könnte den Rückgang bereits verdeutlichen.

	n_{neu}	n	PD
Cluster 4	$n_{4_{neu}}^{\alpha_A} = 30$	132	106.939
Min. Werte	$\lambda_{\min}^{neu} = 41.4$	182	95.869

Tabelle 5.17.: Vergleichswerte für gefährdetes Cluster zum Zeitpunkt $t = 5$

Zum nächsten Analysezeitpunkt, $t = 6$, wirkt sich der starke Rückgang der Objektzahl des vierten Clusters neben den Kardinalitäten weiter auf die innere Struktur des Clusters aus. Da nur noch verhältnismäßig wenige Objekte durch das Cluster absorbiert werden, treten leichte Schwankungen bzgl. Ausdehnung und -richtung des Clusters auf (vgl. Abbildung 5.11d). Auch wenn mit $n_{4_{neu}}^{\alpha_A} = 13$ Objekte nur noch sehr wenige Objekte neu zu dem Cluster hinzukommen, führt dies noch nicht zu einer Clusterelimination, da vor dieser die minimale Anzahl an Neuzuordnungen aufgrund von $t_{eC}^{\max} = 2$ zweimalig unterschritten werden darf. Auf diese Weise können vorübergehende Schwankungen abgefangen werden. Dennoch wird anhand der Evaluation der Gesamtzahl absorbierter Objekte deutlich, dass eine Elimination in Kürze bevorsteht (vgl. Tabelle 5.18). Da $n_{4_{neu}}^{\alpha_A}$ stark rückläufig ist, wird der minimale Wert in der Folgeperiode voraussichtlich ebenfalls unterschritten werden; ferner ist aufgrund der stetigen Abnahme der Gesamtzahl absorbierter Objekte keine ausreichende Clustergröße mehr zu erwarten, da schon zum Zeitpunkt $t = 6$ mit $96 > 0.5 \cdot 191 = 95.5$ die geforderte Mindestgröße noch existierender Cluster grenzwertig ist. Entsprechend wird wie erwartet zum Zeitpunkt $t = 7$ das vierte Cluster aus der Clusterstruktur eliminert (Abbildung 5.11e). Trotz der Häufung der verbliebenen, nun als Ausreißer geltenden Objekte des vierten Clusters wird kein nach Abschnitt 5.4 relevantes

	n_{neu}	n	PD
Cluster 4	$n_{4_{neu}}^{\alpha^{\mathrm{A}}} = 13$	96	84.524
Min. Werte	$\lambda_{\min}^{neu} = 44.1$	191	91.269

Tabelle 5.18.: Vergleichswerte für gefährdetes Cluster zum Zeitpunkt $t = 6$

Ausreißercluster erkannt. Die freien Objekte beeinflussen das dritte Cluster in seiner Ausdehnung gering; diese Modifkation kann jedoch direkt auf die vorgenommene Clusterelimination zurückgeführt werden. Der Folgezeitpunkt $t = 8$, dargestellt in Abbildung 5.11f, zeigt die nahezu komplette Wiederherstellung der Ausgangsstruktur der drei verbliebenen Cluster, während die Objekte des vierten Clusters aus der Struktur fast verschwunden sind.

5.6. Vereinigung von Clustern

Im Zeitverlauf ist es möglich, dass sich verschiedene Cluster aufeinander zubewegen und sich schließlich zu einem einzigen Cluster vereinigen. Die Relevanz dieser Veränderung lässt sich anhand eines Beispiel aus dem Marketingbereich verdeutlichen: Für ein Unternehmen, das ein differenziertes Marketing betreibt, indem es verschiedene Kundensegmente separat bearbeitet, ist die Kenntnis der genauen Struktur elementar. Nähern sich einzelne Kundensegmente aneinander an, so muss prognostiziert werden, ob sich die getrennte Bearbeitung noch rentiert, oder der Zeitpunkt bestimmt werden, zu dem auf eine Differenzierung der Segmente verzichtet werden und die Bearbeitung der Segmente gemeinsam erfolgen soll. Somit liegt der Fokus auf der Vorhersage des Zeitrahmens, ab dem zwei Segmente nicht länger separiert zu bearbeiten sind. Die Beantwortung dieser Frage ist kontextabhängig: In verschiedenen Bereichen kann es erforderlich sein, einzelne Cluster als getrennt zu betrachten, während in anderen Bereichen eine Unterscheidung der Cluster sich nicht länger als sinnvoll erweist. Dieser Umstand soll anhand des bekannten Iris-Datensatzes (Asuncion und Newman, 2007) und den Eigenschaften *petal length* (Länge des Kronblatts) und *petal width* (Breite des Kronblatts) in Abbildung 5.12 verdeutlicht werden. Im botanischen Kontext ist die Unterscheidung der beiden Irisarten Versicolor und Virginica relevant. Nähme man jedoch an, dass die einzelnen Objekte Kunden repräsentierten, wäre obere Objektgruppe eher als ein Segment zu interpretieren.

Im Folgenden erfolgt zunächst eine Beschreibung des allgemeinen Vorgehens zur Clustervereinigung. Die beschriebene Vorgehensweise eignet sich für verschiedene Bereiche; im Falle der Trennung aus einer theoretischen Sicht sind entsprechend Grenzwerte anzupassen, die im hier fokussierten Marketing-Kontext geringer anzusetzen sind. Im Anschluss wird die Vorgehensweise anhand künstlicher Daten veranschaulicht.

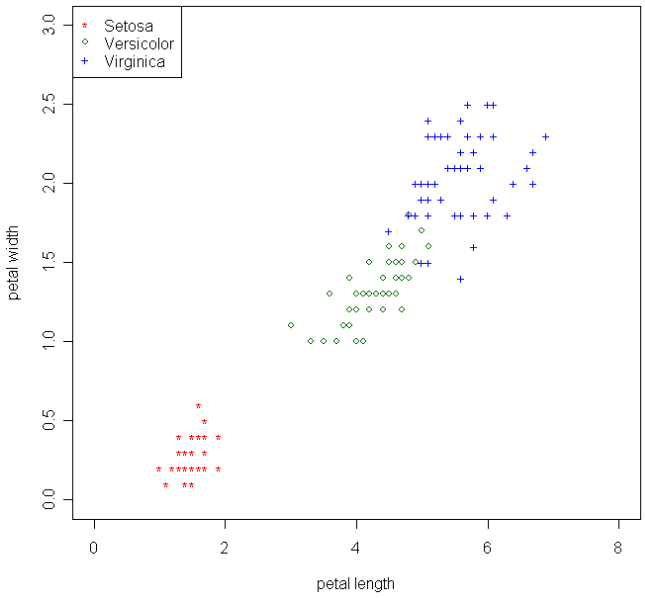


Abbildung 5.12.: Iris-Daten

5.6.1. Vorgehen zur Vereinigung von Clustern

In der Literatur werden verschiedene Algorithmen für das automatische Vereinigen von Clustern vorgestellt, z.B. der zielfunktionsbasierte *Competitive Agglomeration*[22]-Algorithmus von Frigui und Krishnapuram (1997), in dem die verschiedenen Cluster um ihr Bestehen konkurrieren und schwache Cluster eliminiert werden. Das Ziel dieser Algorithmen liegt indessen in der Ermittlung einer geeigneten Clusterzahl für die Gesamtstruktur; nach Beendigung der Analyse besteht keine Möglichkeit, die Ursache möglicher Änderungen nachzuvollziehen bzw. eine Vorhersage zukünftiger Änderungen vorzunehmen. Da im Rahmen des Change Minings der Fokus jedoch darauf liegt, abrupte Veränderungen frühzeitig aufdecken zu können, erweist sich die genannte Form der Clustervereinigung in diesem Kontext als ungeeignet. Es werden hingegen Ansätze benötigt, die auf der Ähnlichkeit und der Kompatibilität verschiedener Cluster beruhen, d.h., die Kriterien zur Bestimmung der Clusternähe verwenden, die unabhängig vom zugrundeliegenden Algorithmus angewandt werden können und nur auf dem Ergebnis der Neuzuordnung von Objekten beruhen. Bei der Bestimmung der Clusterähnlichkeit muss ferner neben der allgemeinen Ähnlichkeit von Clustern evaluiert werden, inwiefern sich die Ähnlichkeit zwischen den Clustern gegenüber den Vorperioden verändert hat, um daraus Informationen und Trends für die zukünftige Entwicklung ziehen zu können.

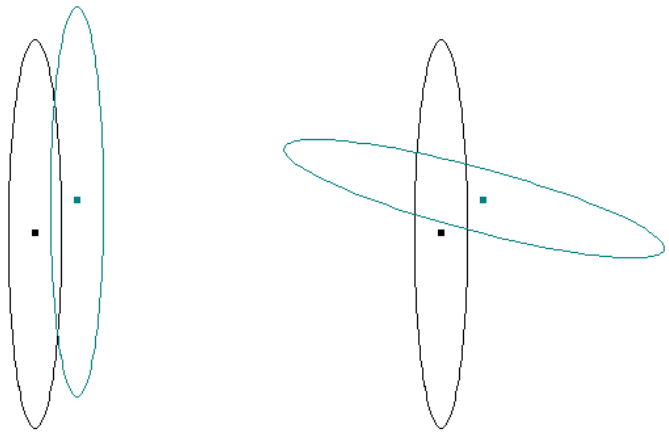
Zum Vereinigen von zwei Clustern C_i und $C_{i'}$ müssen zwei wesentliche Kriterien erfüllt sein (vgl. Krishnapuram und Freg, 1992):

1. Kleine Distanz zwischen den Clusterzentren:

$$\|\vec{v}_i - \vec{v}_{i'}\|^2 = k^1_{ii'} \leq \lambda^V_{dist},\ \lambda^V_{dist} \text{ nahe } 0 \tag{5.29}$$

2. Ausreichende Parallelität der Cluster zueinander, d.h., die Hyperebenen der Cluster müssen ausreichend parallel sein; die Überrpüfung erfolgt anhand der kleinsten Eigenvektoren

[22] Übersetzt bedeutet der Name *Wetteifernde Verdichtung*.



(a) Ausreichende Parallelität (b) Unzureichende Parallelität

Abbildung 5.13.: Bedeutung der Clusterparallelität

der Kovarianzmatrizen:

$$|\vec{e}_{ip} \cdot \vec{e}_{i'p}| = k^2_{ii'} \geq \lambda^V_{par},\ \lambda^V_{par} \text{ nahe } 1, \tag{5.30}$$

wobei $\vec{e}_{\cdot p}$ jeweils den Eigenvektor des kleinsten Eigenwertes der Kovarianzmatrix repräsentiert.

Werden die vorhandenen Cluster unter Verwendung einer Fuzzy-C-Means-Variante als kugelförmig angenommen, entfällt die zweite Bedingung, da die Parallelität in jedem Fall vorliegt; in diesem Fall reicht die Erfüllung des ersten Kriteriums aus. Bei ellipsoiden Clustern muss hingegen überprüft werden, ob die generelle Ausrichtungen der Cluster zuneinander kompatibel sind. Stimmen diese weitestgehend überein, können die einzelnen Cluster vereinigt werden (Abbildung 5.13a); weichen sie hingegen stark voneinander ab, muss unabhängig von der Distanz der Clusterzentren zueinander auf eine Clustervereinigung verzichtet werden (Abbildung 5.13b). Bei ellipsoiden Clustern sind die beiden Kriterien bzgl. Distanz und Parallelität jedoch nicht separat zu analysieren; vielmehr besteht die Möglichkeit, dass eine sehr gute Erfüllung eines Kriteriums die mangelhafte Erreichung des anderen kompensieren kann (vgl. Kaymak und Babuška, 1995). Um diese Kompensation zu realisieren, werden zwei Fuzzy-Mengen benötigt, die die Bedingungen *nahe 1* bzw. *nahe 0* aus den Bedingungen in (5.29) und (5.30) repräsentieren. Geeignete Zugehörigkeitsfunktionen werden unter Verwendung von Exponentialfunktionen definiert, die einen problemabhängigen Support aufweisen. Die Grenze des Supports wird anhand der durchschnittlichen Werte aller $k^1_{ii'}$ in (5.31) bzw. $k^2_{ii'}$ in (5.32) charakterisiert.

$$\nu_1 = \frac{1}{c(c-1)} \sum_{i=1}^{c} \sum_{\substack{i'=1 \\ i' \neq i}}^{c} k^1_{ii'} \tag{5.31}$$

$$\nu_2 = \frac{1}{c(c-1)} \sum_{i=1}^{c} \sum_{\substack{i'=1 \\ i' \neq i}}^{c} k^2_{ii'} \tag{5.32}$$

Auf diese Weise lassen sich exponentielle Zugehörigkeitsfunktionen der Form

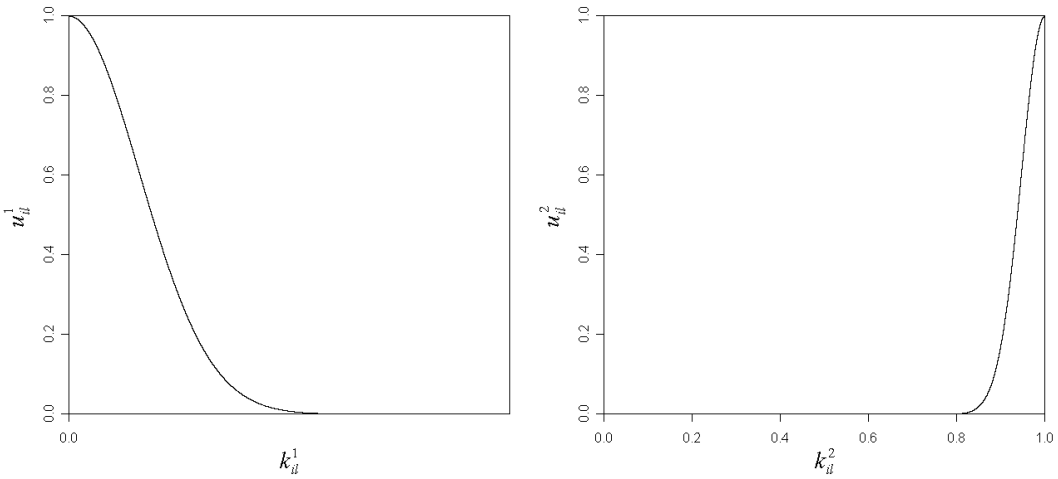
$$u^1_{ii'} = e^{-\frac{K}{\nu_1^2}\left(k^1_{ii'}\right)^2} \tag{5.33}$$

für die Fuzzy-Menge *nahe 0* bzw.

$$u^2_{ii'} = e^{-\frac{K}{(1-\nu_2)^2}\left(1-k^2_{ii'}\right)^2} \tag{5.34}$$

für die Menge *nahe 1* definieren. Dabei ist K eine Konstante, durch die der Support kontextabhängig angepasst werden kann.

Durch die Definition des Supports in Abhängigkeit der generellen Clusterstruktur kann sichergestellt werden, dass die jeweiligen Zugehörigkeitsfunktionen an spezifische Probleme angepasst werden (vgl. Kaymak, 1998, S. 245f.). Abbildung 5.14 zeigt mögliche Zugehörigkeitsfunktionen für die Kriterien.



(a) Distanz nahe 0 (b) Parallelität nahe 1

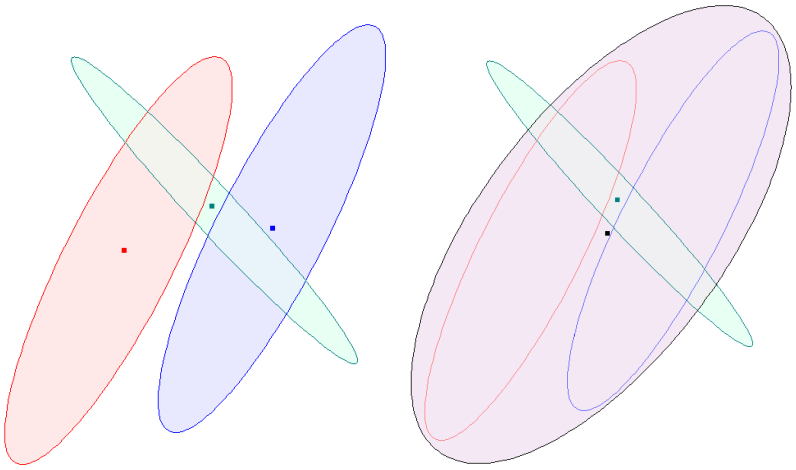
Abbildung 5.14.: Zugehörigkeitsfunktionen für die Fuzzy-Mengen bzgl. Distanz und Parallelität von Clustern

Sind die Zugehörigkeitsgrade zu den Fuzzy-Mengen *nahe 0* und *nahe 1* bekannt, können diese mit Hilfe des Aggregationsoperators

$$s_{ii'} = \left(\frac{(u^1_{ii'})^q + (u^2_{ii'})^q}{2}\right)^{\frac{1}{q}}, q \in \mathbb{R}, \tag{5.35}$$

aggregiert werden, um die Ähnlichkeit unter Einbeziehung der Kompensierung der beiden Kriterien der Cluster C_i und $C_{i'}$ zu bestimmen. Kaymak (1998, S. 250f.) empfiehlt dabei, q= 0.5 zu wählen; empirisch liefere diese Wahl zufriedenstellende Ergebnisse. Die in (5.35) gegebene Aggregation der Zugehörigkeitsgrade unterstellt, dass Distanz und Parallelität in jedem Fall in ihrer Relevanz gleichzusetzen sind. Dies ist jedoch kontextabhängig: So kann in der Informationstheorie die Ausrichtung der Cluster zueinander von übergeordneter Bedeutung sein, d.h., haben zwei Cluster ein nahezu identisches Zentrum, stehen im Extremfall jedoch senkrecht zueinander, so dürfen sie nicht vereinigt werden (vgl. Abbildung 5.13b). Im Marketingkontext hingegen spielt bei der differenzierten Bearbeitung einzelner Segmente die Distanz die entscheidende Rolle, da auch für zueinander senkrecht stehende Segmente eine gemeinsame Bearbeitung erfolgen kann. Daraus ergibt sich eine gewichtete Form der Kompensierung aus (5.35) mit

$$s_{ii'} = \left(\gamma^s \left(u^1_{ii'}\right)^q + (1-\gamma^s)\left(u^2_{ii'}\right)^q\right)^{\frac{1}{q}}, q \in \mathbb{R}, \tag{5.36}$$



(a) Separierte parallele Cluster (b) Vereinigte parallele Cluster

Abbildung 5.15.: Vereinigung bei dazwischen liegendem inkompatiblem Cluster

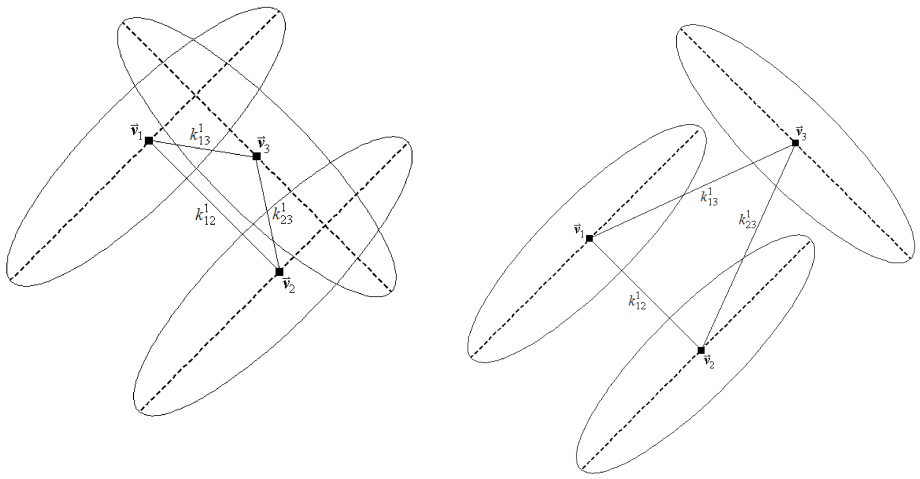
wobei $\gamma^s \in [0,1]$ die Gewichtung der einzelnen Eigenschaften angibt.

In verschiedenen Szenarien mit ellipsoiden Clustern kann es erforderlich sein, dass eine weitere Bedingung zu den oben genannten hinzugefügt werden muss, um eine übermäßige Vereinigung auf Kosten relevanter Informationen zu vermeiden (vgl. Angstenberger, 2000, S. 104f.; Kaymak und Babuška, 1995). Abbildung 5.15a zeigt zwei parallele Cluster, die sehr nah beieinander liegen und unter Umständen vereinigt werden könnten, läge nicht ein weiteres Cluster dazwischen, das sich zu den parallelen Clustern als inkompatibel darstellt. Falls eine Vereinigung der parallelen Cluster vorgenommen würde, wäre der Verlust sämtlicher, das dritte Cluster betreffender Informationen die Folge, da es nicht länger von den anderen Clustern unterschieden werden könnte (Abbildung 5.15b). Liegt demzufolge ein inkompatibles Cluster zwischen den zu vereinigenden Clustern, so ist eine zusätzliche, eine Clustervereinigung verhindernde Bedingung erforderlich:

$$\min \max_{\substack{\vec{v}_i \in Q \\ \vec{v}_k \notin Q}} k^1_{ik} > \max_{\vec{v}_i, \vec{v}_{i'} \in Q} k^1_{ii'}, \tag{5.37}$$

wobei Q die Menge der kompatiblen Cluster angibt. Unter Anwendung der zusätzlichen Bedingung in (5.37) werden lediglich solche Cluster vereinigt, bei denen die maximale Distanz zwischen den zu vereinigenden Zentren kleiner ist als das vergleichbare Minimum der maximalen Distanzen zu allen nicht kompatiblen Zentren, d.h., eine Vereinigung erfolgt nur dann, wenn sich kein weiteres Cluster zwischen den zu vereinigenden Clustern befindet (vgl. Abbildung 5.16).

Wird im Rahmen der Marktforschung eine Marktstruktur untersucht, kann die in (5.37) eingeführte Bedingung vernachlässigt werden, da weniger die Lage der Clusterprototypen zueinander von Bedeutung ist als vielmehr ihr Überschneidungsgrad. So erscheint es wenig plausibel, Marktsegmente getrennt zu bearbeiten, weil sich zwischen ihnen ein weiteres Segment ohne Nachweis einer ausreichenden Kompatiblität befindet; sofern eine gemeinsame Bearbeitung wirtschaftlich als sinnvoll erachtet wird, erfolgt diese trotz des Vorhandenseins eines aufgrund seiner Ausrichtung inkompatiblen Segments. Um diesem Umstand gerecht zu werden, erscheint



(a) Verletzung der Bedingung (b) Keine Verletzung der Bedingung

Abbildung 5.16.: Zusatzbedingung bei ellipsoiden Clustern (vgl. Angstenberger, 2000, S. 104)

es sinnvoll, den Überschneidungsgrad einzelner Cluster separat zu evaluieren. Setnes und Kaymak (1998) führen hierzu ein Inklusionsmaß $\mathtt{I}_{ii'}$ ein, das von Angstenberger (2000, S. 99) für die Verwendung von α-Schnitten angepasst wurde, so dass es für den vorgestellten Ansatz zur Absorbierung geeignet ist:

$$\mathtt{I}_{ii'} = \frac{\operatorname{card}\left([U_i]_{\alpha^{\mathrm{A}}} \cap [U_{i'}]_{\alpha^{\mathrm{A}}}\right)}{\operatorname{card}\left([U_i]_{\alpha^{\mathrm{A}}}\right)}. \tag{5.38}$$

Da es sich bei dem Inklusionsmaß in (5.38) um ein asymmetrisches Maß handelt, wird die Ähnlichkeit zwischen Clustern basierend auf ihrem Überlappungsgrad mit Hilfe von

$$s^{\mathtt{I}}_{ii'} = \max\left\{\mathtt{I}_{ii'}, \mathtt{I}_{i'i}\right\} \tag{5.39}$$

bestimmt.

Um eine Vereinigung ähnlicher Cluster vornehmen zu können, werden die drei Kriterien bzgl. Überschneidungsgrad, Parallelität und Distanz der Clusterzentren in der Reihenfolge ihrer Priorität evaluiert. Zunächst erfolgt eine Prüfung der Ähnlichkeit basierend auf (5.39):

- Gilt $s^{\mathtt{I}}_{ii'} \geq \lambda_V^{\mathtt{I}^{\max}}$, $\lambda_V^{\mathtt{I}^{\max}} \in (0,1]$, erfolgt eine Clustervereinigung ohne Überprüfung der weiteren Kriterien bzgl. Distanz und Parallelität.
- Gilt $\lambda_V^{\mathtt{I}^{\min}} \leq s^{\mathtt{I}}_{ii'} < \lambda_V^{\mathtt{I}^{\max}}$, $\lambda_V^{\mathtt{I}^{\min}} \in [0, \lambda_V^{\mathtt{I}^{\max}}]$, müssen für eine Vereinigung neben der Überschneidung die Distanz und die Parallelität der Cluster zueinander evaluiert werden.
- Gilt $s^{\mathtt{I}}_{ii'} < \lambda_V^{\mathtt{I}^{\min}}$, ist unabhängig von der sonstigen Lage der Cluster zueinander keine Clustervereinigung möglich.

Nur für den Fall, dass $\lambda_V^{\mathtt{I}^{\min}} \leq s^{\mathtt{I}}_{ii'} < \lambda_V^{\mathtt{I}^{\max}}$ gilt, erfolgt eine Überprüfung der Distanz und der Parallelität der Cluster gemäß (5.36), so dass sich die Kriterien gegenseitig kompensieren können. Analog zu den zuvor beschriebenen abrupten Veränderungen müssen auch hier drei Fälle unterschieden werden:

1. Gilt $s_{ii'} \geq \lambda^s_{\max}$, $\lambda^s_{\max} \in (0,1]$, werden die Cluster vereinigt.
2. Gilt $\lambda^s_{\min} \leq s_{ii'} < \lambda^s_{\max}$, $\lambda^s_{\min} \in [0, \lambda^s_{\max}]$, erfolgt keine Vereinigung. Die Cluster werden jedoch vorgemerkt, da eine zukünftige Vereinigung möglich scheint.
3. Gilt $s_{ii'} < \lambda^s_{\min}$, können die Cluster aufgrund ihrer Lage zueinander nicht vereinigt werden.

Sollen alle Cluster, die eine Ähnlichkeit bzgl. des Inklusionsmaßes von mindestens $\lambda_V^{\text{I}^{\min}}$ aufweisen, für eine mögliche zukünftige Vereinigung vorgemerkt werden, wird $\lambda^s_{\min} = 0$ gewählt. Dieses Vorgehen erscheint gerade im Marketingkontext sinnvoll, da die Überlappung einzelner Cluster als wichtigstes Kriterium gilt. Allgemein erhält der Überschneidungsgrad der Cluster bei der vorgestellten Vorgehensweise das höchste Gewicht; nur wenn dieser sich in einer Grauzone befindet, in der keine eindeutige Aussage bzgl. der möglichen Vereinigung von Clustern getroffen werden kann, erfolgt eine Einbeziehung der übrigen Maße. Dieses Vorgehen bietet den Vorteil, dass weiterhin eine Komponsation zwischen Parallelität und Distanz gemäß (5.36) vorgenommen werden kann. Ferner können auf Basis der verschiedenen Kriterien graduelle Veränderungen aufgedeckt werden, die zukünftige Veränderungen implizieren; so kann z.B. aus einer wachsenden Überschneidung eine mögliche Annäherung von Clustern frühzeitig erkannt werden.

Zur Initialisierung einer erneuten Fuzzy-Clusteranalyse wird nach der Vereinigung zweier Cluster C_i und $C_{i'}$ ein gemeinsamer Zugehörigkeitsvektor $\vec{u}_{i\cup i'}$ benötigt. Dieser kann entweder zufällig erzeugt, was jedoch dem Grundsatz der prädiktiven Analyse widerspricht, oder auf Basis der vorhandenen Zugehörigkeitsvektoren $\vec{u}_i$ und $\vec{u}_{i'}$ bestimmt werden. Stutz (1998) empfiehlt, eine einfache Addition der Zugehörigkeitsgrade durchzuführen, d.h.

$$\vec{u}_{i\cup i'} = \vec{u}_i + \vec{u}_{i'}. \tag{5.40}$$

Durch dieses Vorgehen sind theoretisch Zugehörigkeitsgrade möglich, für die $u_{i\cup i',j} > 1$ gilt. Dieses Problem ist jedoch zu vernachlässigen, da die neuen Zugehörigkeitsgrade während der Clusteranalyse nach der Updateregel in (3.5) berechnet werden, so dass bereits nach einer Iteration wieder alle Zugehörigkeitsgrade im zulässigen Bereich liegen.

5.6.2. Untersuchung der Entwicklung vorgemerkter Cluster

Sind Cluster für eine mögliche zukünftige Vereinigung vorgemerkt, muss für die Prognose des Zeitpunkts zum Zusammenfassen die Entwicklung der Cluster zueinander überwacht werden. Dies erfolgt anhand eines Monitorings der Überlappung und Lage der Cluster zueinander über die Zeit hinweg. Neben einer allgemeinen Analyse der Veränderungen der einzelnen Maße zu Inklusion, Distanz und Parallelität sowie der Ähnlichkeit basierend auf der Kompensation von Parallelität und Distanz sind hierbei die Clustertrajektorien elementar, die im Rahmen der Analyse gradueller Unterschiede (vgl. Abschnitt 5.3) bestimmt werden, da diese Aufschluss über die Entwicklung der Position der einzelnen Cluster geben können. Bewegen sich zwei Cluster im Betrachtungszeitraum kontinuierlich aufeinander zu, kann neben der generellen Entwicklung und damit der zeitlichen Vorhersage des Aufeinandertreffens auch die spezifische Veränderung bzgl. der Lage der einzelnen Cluster verfolgt werden, so dass zusätzlich die Position der einzelnen Cluster zum Zeitpunkt der Vereinigung vorhergesagt werden kann. Auf diese Weise ist

es möglich, die relevanten Eigenschaften des neuen Clusterprototypen zu prognostizieren. Die Prognose erfolgt dabei in zwei Schritten:

1. Vorhersage des Zeitpunkts der Vereinigung: Mit Hilfe eines geeigneten Regressionsmodells wird der Zeitpunkt $\hat{t}^V_{ii'}$ vorhergesagt, zu dem entweder die Ähnlichkeit auf Basis des Inklusionsmaßes aus (5.39) ausreicht oder aber für die Kompensation von Distanz und Parallelität nach (5.36) ein genügender Wert zu erwarten ist. Letztere Analyse sollte durch getrennte Betrachtung der einzelnen Maße zu Distanz und Parallelität erfolgen, da bei einer gemeinsamen Betrachtung anhand der Kompensierungsfunktion aus (5.36) wichtige Informationen zur separaten Entwicklung der einzelnen Werte vernachlässigt werden.
2. Prognose der Lage der Cluster C_i und $C_{i'}$ zum Zeitpunkt $\hat{t}^V_{ii'}$ sowie Bestimmung des gemeinsamen Prototypen auf Basis der getrennten Entwicklung.

Auf diese Weise ist es möglich, die zu erwartende Entwicklung bzgl. der Vereinigung zweier Cluster in die Planung einzubeziehen.

5.6.3. Veranschaulichung des Vorgehens anhand künstlicher Daten

Das Vorgehen zum Vereinigen von Clustern wird im Folgenden anhand eines vereinfachten Beispiels verdeutlicht, basierend auf einem zweidimensionalen Datensatz normalverteilter Daten, der anfangs aus vier Clustern bestand. Dabei erhielten zwei der Cluster je Periode konstant 100 Objekte, den anderen wurden jeweils 75 hinzugefügt. Die letztgenannten wurden im Verlauf des Beispiels vereinigt, weshalb sie zunächst auf einer geringeren Objektzahl aufbauten. Insgesamt erstreckt sich das Beispiel über sieben Perioden. Die Zeitfensterlänge wurde analog zu den vorigen Beispielen auf $\tau = 3$ gesetzt, als Analysefrequenz wurde $\Delta t = 1$ gewählt. Die Ausgangswerte der Clusterzentren und der Kovarianzmatrizen zur Generierung der Daten sind in Tabelle 5.19 gegeben; Cluster 2 und Cluster 4 lagen im Vergleich zu den Beispielen zur Neuentstehung bzw. zur Eliminierung eines Clusters näher aneinander, um den Vereinigungsprozess zu beschleunigen. Für Cluster 4 wurde eine konstante Entwicklung um $\begin{pmatrix} 0.3 \\ 0 \end{pmatrix}$ unterstellt, d.h., in jeder Periode änderte sich der Clustermittelpunkt um den angegebenen Vektor, so dass sich die Distanz zu Cluster 2 stetig verringerte.

Cluster	**Objektzahl je Periode**	**Zentrum**	**Kovarianzmatrix**
Cluster 1	100	$\begin{pmatrix} 2 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$
Cluster 2	75	$\begin{pmatrix} 8 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 1.5 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$
Cluster 3	100	$\begin{pmatrix} 12 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 0.8 & -0.5 \\ -0.5 & 0.6 \end{pmatrix}$
Cluster 4	75	$\begin{pmatrix} 3.5 \\ 13 \end{pmatrix}$	$\begin{pmatrix} 1.0 & -0.2 \\ -0.2 & 0.4 \end{pmatrix}$

Tabelle 5.19.: Vorgegebene Parameter

Um den Einfluss des sich ändernden Clusters auf die übrigen Cluster zu reduzieren, wurde zur Fuzzy-Clusteranalyse der Gustafson-Kessel-Algorithmus für ellipsoide Cluster unter Berücksichtigung der Dreiecksbeziehungen der Distanz mit einer Neuberechnung der η_i angewandt (vgl. Abschnitt 4.3.3, Algorithmus 4.7). Für den α-Schnitt wurde entsprechend der vorigen Beispiele $\alpha = 0.5$ gewählt, der Grenzwert der Absorbierung wurde auf $\alpha^{\mathrm{A}} = 0.125$ gesetzt. Ab einem Überlappungsgrad von $\lambda_V^{\mathrm{I}^{\max}} = 0.3$ sollte unabhängig von Distanz und Parallelität zweier Cluster eine Vereinigung erfolgen. Als Grenzwert für den minimalen Überschneidungsgrad, ab dem eine Prüfung bzgl. Distanz und Parallelität zweier Cluster erfolgt, wurde $\lambda_V^{\mathrm{I}^{\min}} = 0.01$ bestimmt, um bereits frühzeitig eine Überlappung aufzudecken. Für ein frühes Vormerken potentiell zu vereinigender Cluster wurde als Grenzwert für die minimale Ähnlichkeit zwischen zwei Clustern nach Aggregation ihrer Distanz und ihrer Parallelität $\lambda^s_{\min} = 0.1$ gesetzt, ab einem Ähnlichkeitswert von $\lambda^s_{\max} = 0.75$ wurden die Cluster vereinigt. Zur Berechnung der Zugehörigkeitsgrade zu den Fuzzy-Mengen *nahe 0* bzw. *nahe 1* wurde in (5.33) und (5.34) $K = 7$ gewählt. Aufgrund der höheren Bedeutung der Distanz zweier Cluster gegenüber ihrer Parallelität im Marketing-Kontext erhielt diese mit $\gamma^s = 0.7$ in (5.36) bei der Aggregation ein höheres Gewicht.

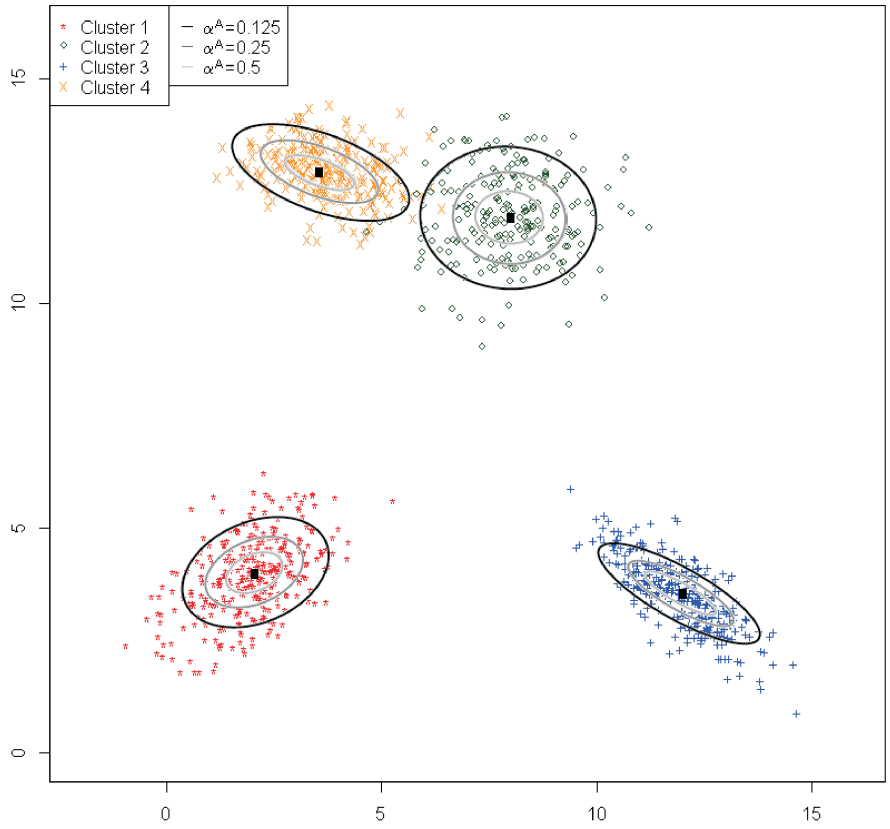
In Abbildung 5.17 sind die Analyseergebnisse grafisch dargestellt. Zum Zeitpunkt $t = 3$ werden wie erwartet vier voneinander getrennte Cluster aufgedeckt (vgl. Abbildung 5.17a). Cluster 2 und Cluster 4 liegen bereits verhältnismäßig nah zueinander, zeigen jedoch noch keinerlei Überlappung und ermöglichen somit keine Vereinigung. Schon zum nächsten Analysezeitpunkt, dargestellt in Abbildung 5.17b, wird die Annäherung der Cluster deutlich: Während die übrigen Cluster nahezu unverändert sind, zeigen Cluster 2 und Cluster 4 eine Überlappung. Diese ist mit einem Ähnlichkeitswert von $s^{\mathrm{I}}_{24} = 0.00659 < 0.01$ jedoch noch zu gering, so dass an dieser Stelle keine weitere Prüfung der Cluster bzgl. ihrer Distanz und ihrer Parallelität erfolgt; diese wäre nur bei der Wahl eines geringeren Wertes für $\lambda_V^{\mathrm{I}^{\min}}$ möglich.

Zum Zeitpunkt $t = 5$ zeigen Cluster 2 und Cluster 4 eine deutliche Veränderung im Verhältnis zueinander (vgl. Abbildung 5.17c): Die Überschneidung der Cluster wächst zunehmend. Als Folge dieses Geschehens wird die Bildung der Kovarianzmatrizen der einzelnen Cluster durch die zusätzlich auftretenden Objekte des jeweils anderen Clusters beeinflusst und weiterhin eine Ausrichtung der die Cluster charakterisierenden Ellipsen in Richtung des anderen Clusters bewirkt. Für Cluster 2 wird eine stärkere Änderung verzeichnet, da die Objekte von Cluster 4 eine vergleichsweise hohe Dichte aufweisen. Trotz der messbaren Annäherung ist aber weiterhin keine Vereinigung der Cluster möglich: Aufgrund des Überschneidungsgrads von $s^{\mathrm{I}}_{24} = 0.01902$ erfolgt zwar eine Prüfung von Distanz und Parallelität, die aus der Kompensation resultierende Ähnlichkeit liegt jedoch mit $s_{24} = 0.67359$ unter dem geforderten Wert von $\lambda^s_{\max} = 0.75$, so dass die Cluster lediglich als potentiell zukünftig zu vereinigen markiert werden; die für die Bestimmung der Ähnlichkeit relevanten Werte sind in Tabelle 5.20 gegeben.

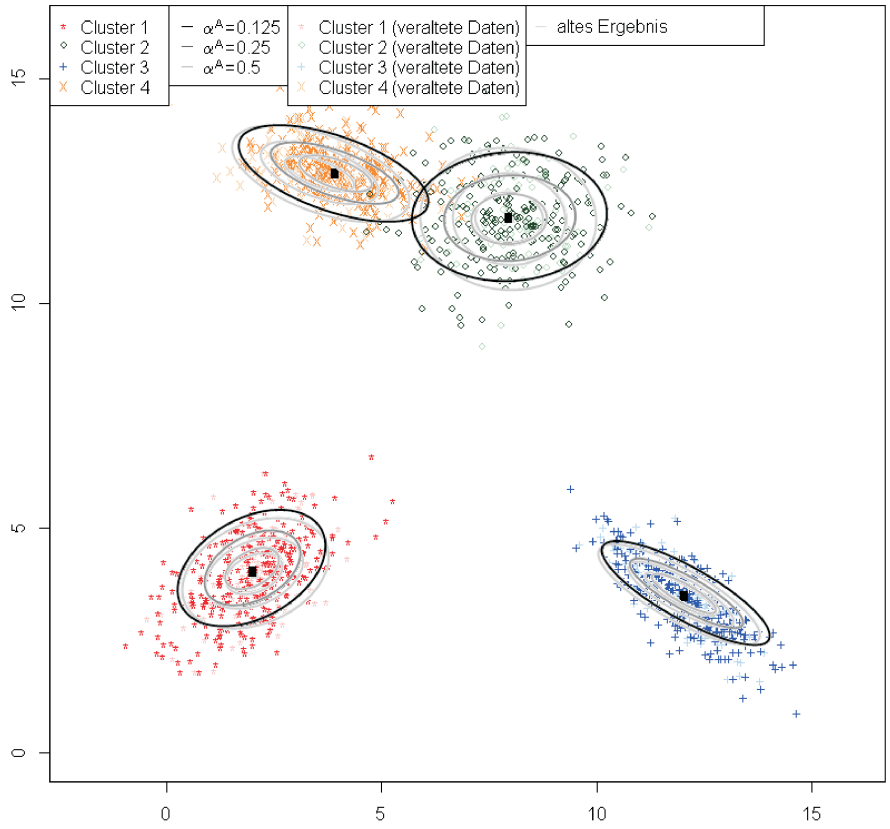
	Distanz	**Parallelität**
Vergleich Cluster 2 und Cluster 4	$k^1_{24} = 17.82313$	$k^2_{24} = 0.87263$
Support für Zugehörigkeitsfunktion	$\nu_1 = 88.97801$	$\nu_2 = 0.59439$

Tabelle 5.20.: Vergleichswerte für potentiell zu vereinigende Cluster zum Zeitpunkt $t = 5$

Da nur für das aktuelle Zeitfenster Werte zu Distanz und Parallelität vorhanden sind, ist noch keine Prognose zur Entwicklung der Cluster zueinander auf Basis eines Regressionsmodells mög-

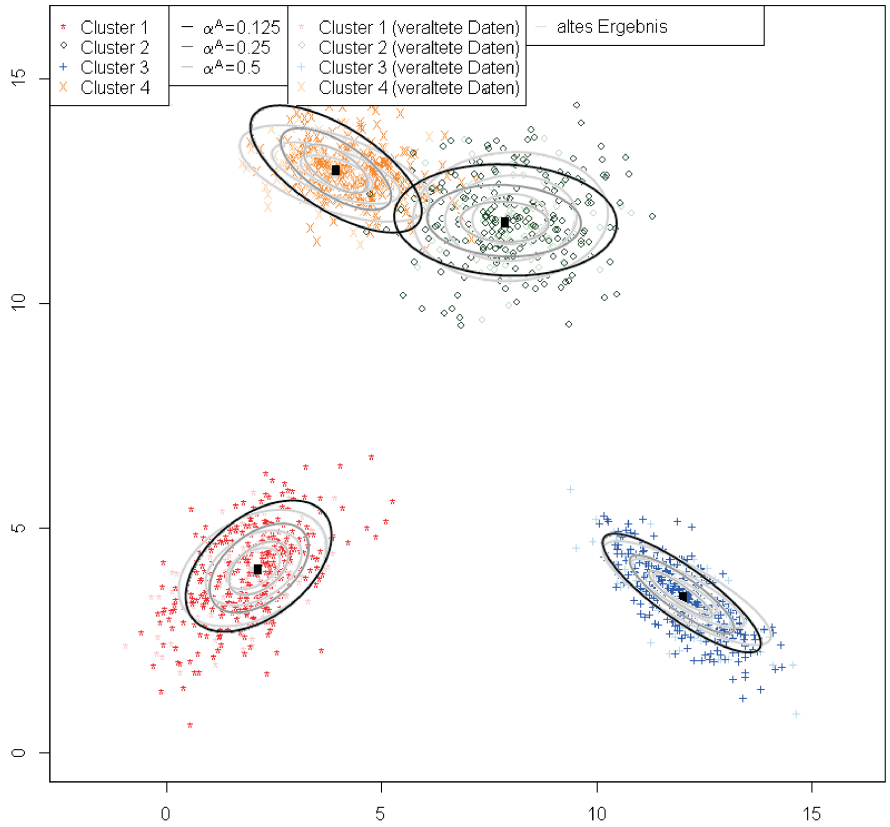


(a) Clusterstruktur für den Zeitpunkt $t = 3$

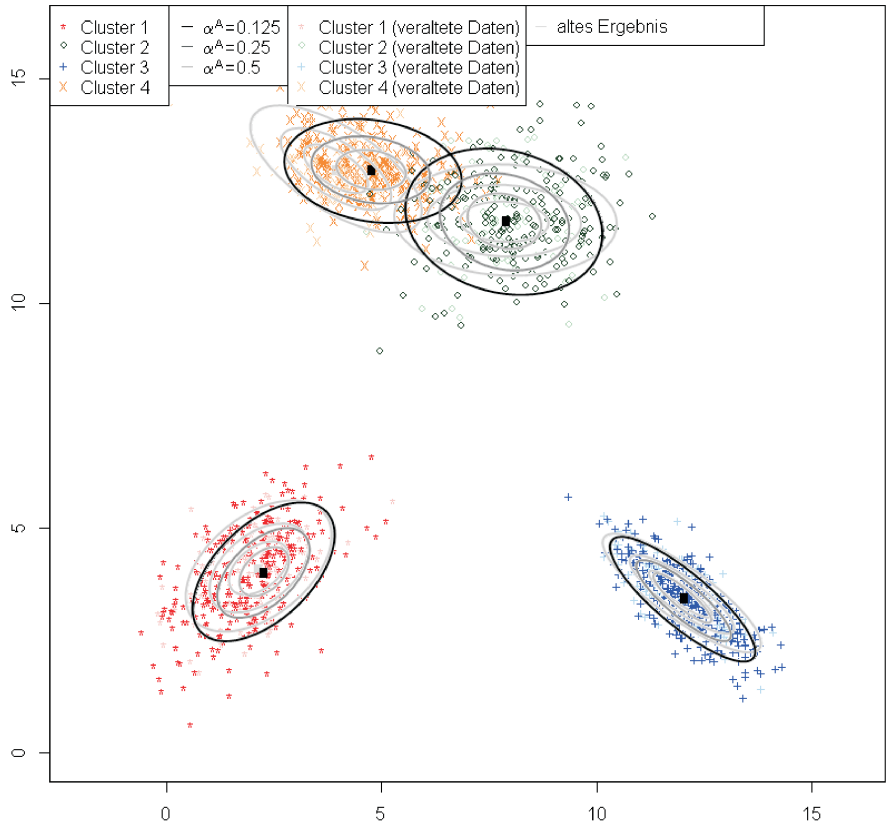


(b) Clusterstruktur für den Zeitpunkt $t = 4$

Abbildung 5.17.: Beispiel für Clustervereinigung

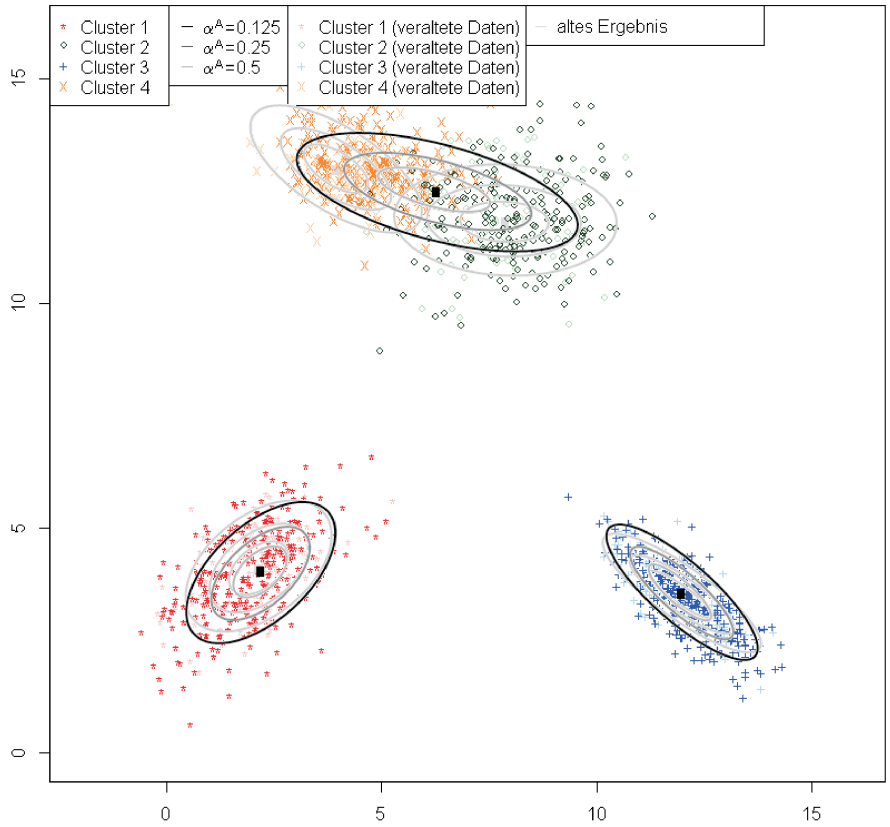


(c) Clusterstruktur für den Zeitpunkt $t = 5$

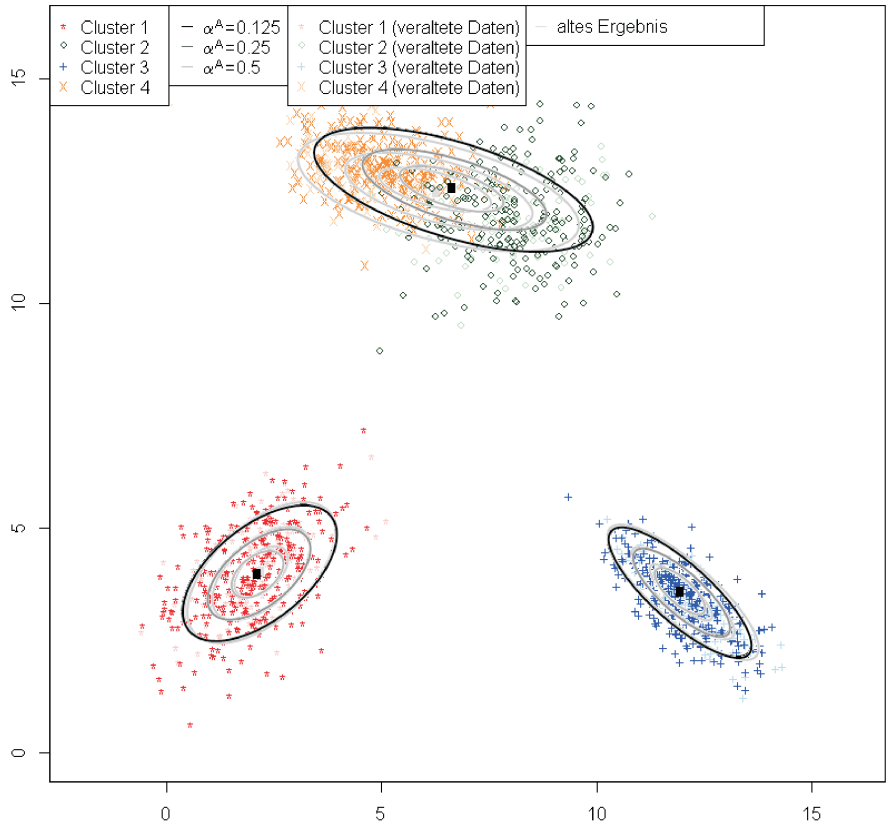


(d) Clusterstruktur für den Zeitpunkt $t = 6$

Abbildung 5.17.: Beispiel für Clustervereinigung



(e) Clusterstruktur für den Zeitpunkt $t = 6$ (nach Reclustering)



(f) Clusterstruktur für den Zeitpunkt $t = 7$

Abbildung 5.17.: Beispiel für Clustervereinigung

lich. Die relativ geringe Distanz bei gleichzeitig hoher Parallelität und der daraus resultierende hohe Ähnlichkeitswert der Cluster deuten jedoch an, dass eine Vereinigung bevorstehen könnte.

Zum Zeitpunkt $t = 6$ erfolgt die erwartete Änderung: Während Cluster 1 und Cluster 3 in Abbildung 5.17d weiterhin keine signifkanten Änderungen aufweisen, ist die Überschneidung zwischen Cluster 2 und Cluster 4 derartig angewachsen, dass im hier fokussierten Marketing-Kontext eine getrennte Bearbeitung zwei entsprechender Segmente nur noch schwer zu begründen wäre. Diese Tatsache wird insbesondere durch die weitere Annäherung der Kovarianzmatrizen bzw. der dargestellten, die Cluster charakterisierenden Ellipsen bestärkt. Aufgrund der geringen Zugehörigkeitsgrade der Objekte in der Schnittmenge der Cluster ist die Ähnlichkeit der Cluster basierend auf ihrer Überlappung mit einem Wert von $s_{24}^{\mathrm{I}} = 0.03265$ zwar immer noch verhältnismäßig gering[23], dennoch werden die Cluster unter Einbeziehung ihrer Distanz und ihrer Parallelität und dem resultierenden Ähnlichkeitswert von $s_{24} = 0.92125$ vereinigt. Die relevanten Werte zur Bestimmung der Ähnlichkeit sind in Tabelle 5.21 gegeben.

	Distanz	**Parallelität**
Vergleich Cluster 2 und Cluster 4	$k_{24}^1 = 11.03398$	$k_{24}^2 = 0.99927$
Support für Zugehörigkeitsfunktion	$\nu_1 = 84.90694$	$\nu_2 = 0.61204$

Tabelle 5.21.: Vergleichswerte für potentiell zu vereinigende Cluster zum Zeitpunkt $t = 6$

Aufgrund der abrupten Veränderung zum Zeitpunkt $t = 6$ ist in Abbildung 5.17e das Ergebnis des Reclusterings gesondert aufgeführt. Nach der erneuten Analyse bleiben Cluster 1 und 3 unbeeinflusst von der aufgetretenen Änderung innerhalb der Clusterstruktur. Für Cluster 2 und Cluster 4 wird nur noch das erwartete gemeinsame Cluster aufgedeckt, dessen Lage und Ausrichtung intuitiv naheliegend erscheinen. Auch zum Folgezeitpunkt $t = 7$, dargestellt in Abbildung 5.17f, werden keine weiteren abrupten Änderungen aufgedeckt, d.h., es werden keine neuen Cluster im Umfeld des vereinigten Clusters aufgezeigt. Lediglich eine leichte Änderung des vereinigten Clusters aufgrund der weiteren Annäherung der Ursprungscluster kann verzeichnet werden, während sich die übrigen Cluster weiterhin nahezu unverändert darstellen.

5.7. Trennen eines Clusters

Die vierte abrupte Veränderung, die innerhalb eines Clusters im Zeitablauf auftreten kann, besteht in der Aufteilung eines Clusters in Teil- bzw. Subcluster, wenn verschiedene Regionen erhöhter Dichte innerhalb des Clusters entstehen und das ursprüngliche Clusterzentrum im

[23] Aufgrund der vorhandenen Ähnlichkeitswerte auf Basis des Inklusionsmaßes nach (5.38) aus den vorherigen Zeitfenstern ist eine Schätzung zur Vorhersage eines ausreichenden Überschneidungsgrades zumindest theoretisch möglich. Die lineare Schätzfunktion liegt dementsprechend bei $\hat{s}_{24}(t) = 0.01243t - 0.04313$ und liefert als aktuellen Schätzwert $\hat{s}_{24}(6) = 0.03145$. Eine solch lineare Schätzung ist jedoch nicht sinnvoll, da bei zunehmender Annäherung der Cluster vermehrt Objekte mit höheren Zugehörigkeitsgraden in der Schnittmenge vorkommen und daher ein überproportionales Wachstum zu erwarten ist. Entsprechend wird die zu erwartende Zeitspanne bis zum Erreichen einer ausreichenden Überlappung wegen des linearen Modells und der zunächst geringen Zuwachsraten stark überschätzt.

Extremfall verwaist. Dies tritt immer dann auf, wenn die verschiedenen durch das Cluster absorbierten Objekte sich in unterschiedliche Richtungen entwickeln; im Marketing-Kontext kann z.B. die Änderung der Interessen auf Kundenseite aufgrund neuer Entwicklungen und Trends zu einer solchen abrupten Veränderung der Struktur führen, d.h. aus einer vorher homogenen Gruppe entwickeln sich heterogene Teilgruppen. Ein Unternehmen muss diese Veränderung zügig erkennen und aufdecken, um weiterhin fähig zu sein, die entstehenden Segmente getrennt zu bearbeiten, sofern dies sinnvoll erscheint.

Wie zuvor beim Vereinigen von Clustern ist auch beim Teilen eines Clusters der Kontext relevant; in einigen Bereichen kann es von Bedeutung sein, bereits früh eine Trennung zu vollziehen. Das in Abschnitt 5.6, Abbildung 5.12 angeführte Beispiel verdeutlicht diesen Umstand. In anderen Bereichen wie der Marktforschung erweist sich eine Trennung erst dann als erforderlich, wenn eine deutliche Separierung der einzelnen Teilcluster vorliegt.

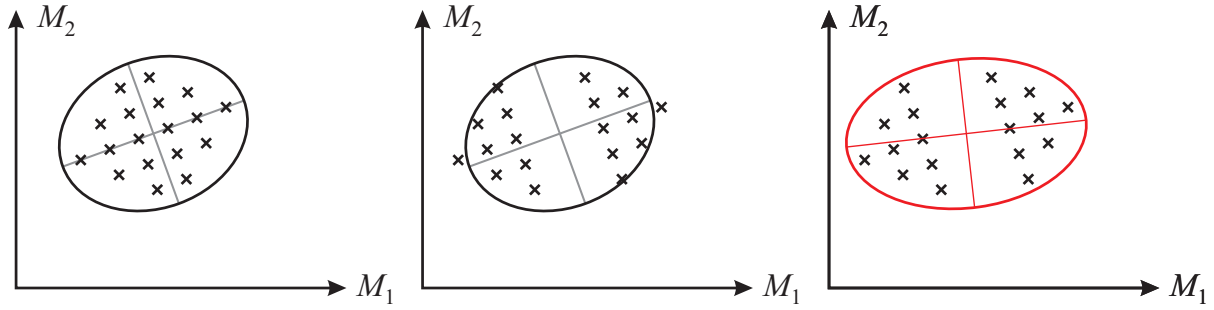
Im Folgenden wird zunächst wieder die allgemeine Vorgehensweise im Detail beschrieben, bevor diese anhand eines künstlichen Beispiels verdeutlicht wird.

5.7.1. Vorgehen zum Trennen eines Clusters

Die komplizierteste Art abrupter Veränderungen betrifft die Trennung eines Clusters. Die Problematik besteht in der Schwierigkeit, frühzeitig eine einheitliche Veränderung innerhalb der Struktur eines einzelnen Clusters nachzuvollziehen, die eine Variation der Dichteverteilung innerhalb des Clusters bewirkt. Eine Umverteilung der Objekte innerhalb eines Clusters wird häufig erst nach einigen Perioden deutlich erkennbar. Diese Art der Veränderung stellt insbesondere dann eine Herausforderung dar, wenn keine festen Objekte mit entsprechenden, eine einheitliche Entwicklung einzelner Objekte anzeigenden Trajektorien vorliegen, sondern nur die Struktur des Clusters als Ganzes analysiert werden kann. Eine Clustertrennung geht in der Regel mit verschiedenen graduellen Änderungen einher (vgl. Abschnitt 5.3): Bei einer isolierten Betrachtung verringert sich i.d.R. die Objektzahl, sofern keine allgemeine Zunahme der Objektzahl innerhalb des Clusters vorhanden ist, während das Volumen zunimmt; auch kann sich die Ausrichtung des Clusters ändern. Abbildung 5.18 verdeutlicht diese Anpassungen anhand eines stark vereinfachten Beispiels. Basierend auf der in Abbildung 5.18a gegebenen Ausgangslage erfolgt die Clustertrennung im Beispiel aufgrund eines einfachen Auseinanderdriftens der Objekte des Clusters in der ersten Dimension. Die Schätzung der Objektzahl auf Basis der alten Clusterstruktur ergibt gemäß Abbildung 5.18b rückläufige Kardinalitäten. Das Ergebnis des Updates der zugehörigen Kovarianzmatrix zeigt außerdem, dass sich die Ausdehnung des Clusters erhöht; gleichzeitig ändert sich die generelle Ausrichtung (vgl. Abbildung 5.18c).

Das wichtigste Kriterium zur Verdeutlichung einer bevorstehenden Clustertrennung besteht in der Dichteänderung innerhalb des Clusters; hierzu erweist sich die Analyse der Partitionsdichte neben der einfachen Verwendung der allgemeineren Fuzzy-Kardinalität als sinnvoll. Die Indikation einer zukünftigen Clusterteilung erscheint dabei zunächst ähnlich der zum Aufdecken gefährdeter bzw. zu eliminierender Cluster (vgl. Abschnitt 5.5). Entscheidend sind die folgenden Eigenschaften:

1. *Rückgang der Fuzzy-Kardinalität innerhalb des durch α^A begrenzten Betrachtungsraums*: Die Umverteilung der Objekte führt zu einer Reduktion der Fuzzy-Kardinalität und im



(a) Ausgangslage (b) Verwaisung des Clusterzentrums (c) Neuausrichtung des Clusters

Abbildung 5.18.: Graduelle Änderungen bei Verwaisung des Clusterzentrums

Speziellen der Partitionsdichte, da eine stark veränderte Dichteverteilung vorliegt. Bei einem Auseinanderdriften analog dem Beispiel in Abbildung 5.18 ist außerdem die absolute Kardinalität rückläufig.

2. *Zunahme des Volumens*:
   Durch die Verwaisung des Clusterzentrums und die Verlagerung der Objekte nach außen bzw. zu anderen Häufungspunkten kann ein steigendes Clustervolumen verzeichnet werden.

Um zu unterscheiden, ob eine Clustertrennung impliziert wird oder eine Clusterelimination bevorsteht, muss eine zusätzliche Prüfung weiterer Maße erfolgen. Im Falle einer bevorstehenden Clustertrennung ist neben einer Änderungen bzgl. der Kardinalitäten und des Volumens sowie der Dichte eine Reduktion der Kompaktheit zu beobachten, da das ursprüngliche Clusterzentrum ausdünnt, während eine Umverteilung der Objekte innerhalb des Clusters stattfindet. Diese lässt sich anhand einer Zunahme des Kompaktheitsindexes $\kappa_i^{\alpha^A}$ nach Bensaid u. a. (1996) in (5.7) messen, die einen Rückgang der allgemeinen Kompaktheit des Cluster impliziert. Ferner ist, wie in Abbildung 5.18 bereits grafisch dargestellt, bei der Betrachtung ellipsoider Cluster im Gegensatz zu einer bevorstehenden Clusterelimination aufgrund der Verlagerung der Objekte in der Regel eine Änderung der Ausrichtung zu verzeichnen; dies gilt jedoch nicht, wenn die Änderung entlang der Ausrichtungsachse erfolgt. Es bleibt aber auch im Ausnahmefall eine Änderung der Verhältnisse der Eigenwerte aufgrund der Dehnung in einzelne Richtungen festzustellen. Die genannten Änderungen können jedoch ebenso durch das gemeinsame Auftreten einer einfachen Volumenerhöhung bei gleichzeitiger Clusterdrehung auftreten (vgl. Abschnitt 5.3.2, Tabelle 5.7) und erfordern deswegen eine weitere Differenzierung. Bei einer Volumenerhöhung gilt, dass der Kompaktheitsindex nur bei relativ kleinen Werten für den Grenzwert der Absorbierung α^A eine signifikante Änderung zeigt. Tritt hingegen eine Dichteänderung mit Bildung verschiedener Häufungspunkte innerhalb der clustereigenen Struktur auf, ändert sich auch die Kompaktheit im Clusterkern, d.h. bei vergleichsweise hohen Werten für α^A. Dies ist auf die unterschiedlichen, durch α^A begrenzten Betrachtungsräume zurückzuführen: Im Falle einer Umverteilung und damit einhergehend einer Verwaisung des Clusterzentrums befinden sich bei größeren Werten für den Absorptionsgrenzwert weniger Objekte im betrachteten Bereich; der Rückgang der Kompaktheit trotz Normierung über die Fuzzy-Kardinalität wird dadurch offensichtlich. Bei einer einfachen Volumenerhöhung hingegen wird für einen höheren Grenzwert der Absorbierung ein kompakterer Bereich betrachtet, so dass sich die Änderung des Kom-

paktheitsindexes weniger deutlich zeigt. Gleiches gilt für die Kardinalitäten. Da jedoch der Kompaktheitsindex laut Tabelle 5.7 in Abschnitt 5.3.2 das entscheidende Erkennungsmerkmal für Volumenänderungen darstellt, muss hier eine gesonderte Auswertung erfolgen. Entsprechend wird ein weiterer Grenzwert $\alpha_\kappa^{\mathrm{A}} \in (\alpha^{\mathrm{A}}, 1]$ benötigt, so dass eine Differenzierung zwischen einer bevorstehenden Clusterteilung und einer Volumenerhöhung bei gleichzeitiger Clusterdrehung ermöglicht wird.

Zusammenfassend sind neben der allgemeinen Änderung der Kardinalitäten und des Clustervolumens entsprechend der Überprüfung gefährdeter Cluster in Abschnitt 5.5 folgende Kriterien zu überprüfen:

1. Zunahme der Kompaktheitsindizes bzgl. α^{A} sowie $\alpha_\kappa^{\mathrm{A}}$: Nur wenn für beide Grenzwerte der Absorbierung signifikante Zunahmen bzgl. der Kompaktheitsindizes $\kappa_i^{\alpha^{\mathrm{A}}}$ bzw. $\kappa_i^{\alpha_\kappa^{\mathrm{A}}}$ zu verzeichnen sind, besteht die Möglichkeit einer bevorstehenden Clusterteilung. Ansonsten ist das Cluster als gefährdet zu betrachten und im Rahmen der Clusterelimination weiter zu untersuchen.
2. Änderung der Clusterdehnung bzw. -ausrichtung bei ellipsoiden Clustern: Gilt

$$\left|\vec{e}_{ip}^{\,t-\Delta t} \cdot \vec{e}_{ip}^{\,t}\right| \leq \lambda_{par}^{CT}, \tag{5.41}$$

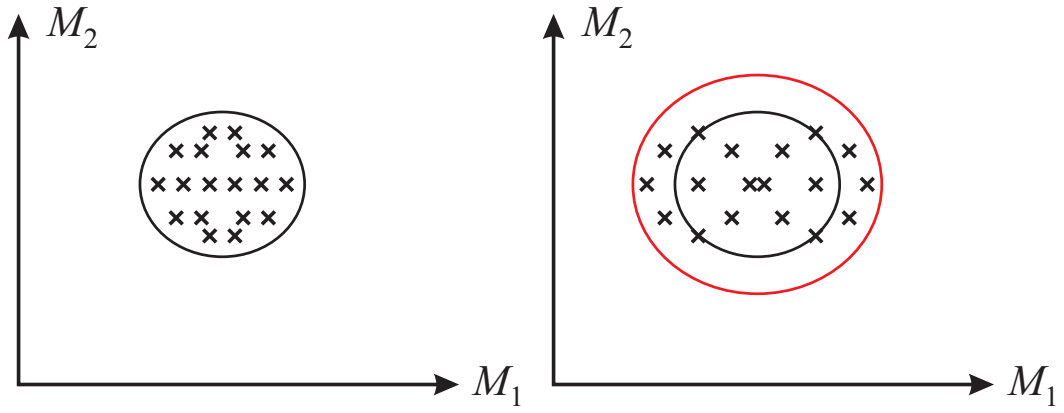
wobei $\lambda_{par}^{CT} \in [0, 1]$ die minimal benötigte Clusterdrehung unter Überprüfung der Eigenvektoren des jeweils kleinsten Eigenwertes angibt (vgl. (5.30) in Abschnitt 5.6), oder

$$\left|\frac{\theta_{ip}^{t}}{\theta_{i1}^{t}} - \frac{\theta_{ip}^{t-\Delta t}}{\theta_{i1}^{t-\Delta t}}\right| \geq \lambda_\theta^{CT}, \tag{5.42}$$

wobei $\lambda_\theta^{CT} > 0$ die minimal geforderte Ausdehnungsänderung festlegt, ist eine Clustertrennung zumindest zukünftig möglich.

Bei der Überprüfung der genannten Kriterien bleibt zu beachten, dass nicht immer der Vergleich mit der vorherigen Periode ausreicht. Um auch langsame, eher schleichend verlaufende Änderungen aufdecken zu können, ist ein längerfristiges Monitoring der Werte über mehrere Analysezeitpunkte vorzunehmen, um so einer Vernachlässigung eventuell möglicher Trennungen aufgrund ihres zögerlichen Verlaufs vorzubeugen. Ferner muss im Falle der Evaluation der Änderungen bzgl. Clusterausrichtung und -ausdehnung beachtet werden, dass nicht beide Änderungen gleichzeitig auftreten müssen, es vielmehr ausreicht, wenn eine der genannten Änderungen zu verzeichnen ist. Auf diese Weise wird ausgeschlossen, dass der zuvor beschriebene Ausnahmefall zur Änderung entlang der Ausrichtungsachse missachtet wird.

Sind die Kriterien für eine mögliche Clusterteilung erfüllt, muss überprüft werden, ob die Trennung des Clusters bereits stattgefunden hat oder nur die Möglichkeit einer zukünftigen Trennung besteht. Hierzu wird eine Clusteranalyse aller dem Cluster zugeordneten Objekte durchgeführt, in der immer genau eine Teilung in einem Analyseschritt vorgenommen wird. Anschließend können die resultierenden Cluster hinsichtlich einer weiteren potentiellen Trennung untersucht werden, so dass eine iterative Zerlegung in mehrere Cluster möglich ist. Alternativ kann vorab eine geeignete Clusterzahl bestimmt werden. Dieses Vorgehen erscheint jedoch aufgrund des Aufwands nicht sinnvoll, da ein Cluster i.d.R. nur in sehr wenige Subcluster zerlegt werden kann.



(a) Ausgangslage (b) Clusterteilung

Abbildung 5.19.: Unterschiedliche Betrachtungsräume

Für die Einteilung eines Clusters in Subcluster, im Folgenden als *Subclustering* bezeichnet, wird ein Wert $\alpha_{SC}^{\mathrm{A}} \in [0, \alpha^{\mathrm{A}})$ benötigt, der angibt, ab welchem Zugehörigkeitsgrad zum Ausgangscluster Objekte in die Analyse einbezogen werden sollen. Dieses Vorgehen ist notwendig, da beim Auseinanderdriften eines Clusters Objekte den durch α^{A} begrenzten Bereich verlassen, jedoch für das Subclustering von Bedeutung sein können, wie Abbildung 5.19 zeigt. Für die Analyse der graduellen Änderungen reicht der kleinere Betrachtungsraum auf Basis der Ausgangslage in Abbildung 5.19a aus. Da durch die Clusterteilung jedoch einige für die Darstellung der Clusterstruktur relevante Objekte diesen Betrachtungsraum verlassen haben, wird für ein anschließendes Subclustering ein entsprechend größerer Betrachtungsraum benötigt (vgl. Abbildung 5.19b), der durch den Wert $\alpha_{SC}^{\mathrm{A}} < \alpha^{\mathrm{A}}$ repräsentiert wird.

Auch wenn beim Subclustering die Clusteranziehung der klassischen possibilistischen Fuzzy-Clusteranalyse eine falsche oder zumindest zu frühe Clusterteilung verhindern könnte, ist ihre Anwendung im Gegensatz zur Neubildung von Clustern (vgl. Abschnitt 5.4) nicht zu empfehlen. Wie in Kapitel 4 beschrieben, benötigt der Originalansatz zur possibilistischen Analyse einen gewissen Separationsgrad der Objekte, um eine eindeutige Clustereinteilung vornehmen zu können. Dieser ist jedoch bei der Betrachtung eines einzelnen sich auseinanderentwickelnden Clusters i.d.R. nicht oder aber erst spät gegeben; aus diesem Grund ist beim Subclustering auf Basis der einfachen possiblistischen Clusteranalyse eine Vorhersage zukünftiger Clusterteilungen nur eingeschränkt möglich. Daher sollte auf einen geeigneten Algorithmus zurückgegriffen werden; insbesondere die Algorithmen zur Einbeziehung der Clusterhomogenität (vgl. Abschnitt 4.3) empfehlen sich hierzu aufgrund ihrer Fokussierung auf Häufungspunkte. Allgemein kann jedoch analyseabhängig der Algorithmus verwendet werden, der auch zur generellen Untersuchung der Clusterstruktur herangezogen wird.

Im Anschluss an das Subclustering erfolgt zunächst eine Prüfung, ob es sich bei den gefundenen Clustern tatsächlich um zwei separate Cluster handelt und nicht eines doppelt gefunden wird (vgl. Abschnitt 5.4 zur Clusterneubildung). Diese erfolgt auf Basis der Ähnlichkeit unter Berücksichtigung der Inklusion nach (5.39). Nur bei einer ausreichenden Clusterseparation der Subcluster C_{i_1} und C_{i_2} des Clusters C_i ist eine bereits stattgefundene Teilung möglich, d.h.,

wenn gilt

$$s^{\mathrm{I}}_{i_1 i_2} \leq \lambda^{\mathrm{I}^{\max}}_{CT}, \tag{5.43}$$

wobei $\lambda^{\mathrm{I}^{\max}}_{CT} \in [0,1]$ den Grenzwert für die maximal zulässige Überschneidung zweier Subcluster angibt. Trifft dies zu, wird das Cluster als potentiell teilbar markiert. Der Grenzwert sollte dabei hoch gewählt werden, d.h. $\lambda^{\mathrm{I}^{\max}}_{CT} \in \left[\lambda^{\mathrm{I}^{\max}}_{V}, 1\right]$ (vgl. Abschnitt 5.6.1), so dass ein Cluster, das später getrennt werden kann, frühzeitig markiert wird.

Wird die Bedingung aus (5.43) erfüllt, müssen weitere Kriterien geprüft werden, um festzustellen, ob die Teilung zu übernehmen oder nur vorzumerken ist. Zunächst muss analog zum Vorgehen beim Vereinigen von Clustern evaluiert werden, ob eine so deutliche Separierung der Subcluster vorliegt, um als individuelle Cluster wahrgenommen zu werden. Dies geschieht neben einer Verschärfung der Bedingung (5.43) zu

$$s^{\mathrm{I}}_{i_1 i_2} \leq \lambda^{\mathrm{I}^{\min}}_{CT} \tag{5.44}$$

mit $\lambda^{\mathrm{I}^{\min}}_{CT} \in [0, \lambda^{\mathrm{I}^{\max}}_{CT}]$ anhand der Distanz zwischen den Zentren der Subcluster. Zur Distanzevaluation wird dabei analog zum Vorgehen zur Clustervereinigung die quadrierte euklidische Distanz verwendet. Gilt

$$\|\vec{v}_{i_2} - \vec{v}_{i_1}\|^2 \leq \lambda^{CT}_{dist}, \tag{5.45}$$

wobei λ^{CT}_{dist} entweder kontextabhängig zu wählen oder abhängig von der übrigen Clusterstruktur zu bestimmen ist (vgl. Abschnitt 5.6.1), gelten die Cluster als ausreichend separiert; ansonsten wird das Cluster für eine zukünftige Teilung vorgemerkt. Durch die Verwendung der quadrierten euklidischen Distanz anstelle der Mahalanobisdistanz bei ellipsoiden Clustern wird vermieden, dass aufgrund der Ausrichtung eines einzelnen Clusters die Clusterteilung verhindert wird, obwohl eine eindeutige Unterscheidung möglich ist. Abbildung 5.20 verdeutlicht diesen Umstand: Zöge man die Kovarianzmatrizen in die Bestimmung der Distanz zwischen den Zentren ein, entweder durch Bildung des Mittelwertes der asymmetrischen Clusterdistanzen oder aber durch Verwendung des Minimalwertes oder ähnliches, so fiele ins Gewicht, dass sich das Zentrum des linken Subclusters innerhalb des Absorptionsbereichs des rechten Clusters befindet, dementsprechend keine ausreichende Clustertrennung vorläge. Diese Problematik könnte zwar verhindert werden, indem in jedem Fall der maximale Distanzwert verwendet würde; dies erschwerte jedoch den Vergleich zu den Distanzen in der allgemeinen Clusterstruktur. Außerdem reicht die Verwendung der quadrierten euklidischen Distanz aus, um eine ausreichende Separierung der Clusterzentren zu prüfen.

Liegt eine ausreichende Trennung der Cluster vor, müssen die Subcluster analog der Neubildung von Clustern in Abschnitt 5.4 über eine ausreichende Größe und Dichte im Vergleich zu den übrigen Clustern verfügen. Ihre Evaluation erfolgt in Anlehnung an die in Abschnitt 5.4 unter Einbeziehung der minimalen Objektzahl bzw. Dichte der ungefährdeten Cluster beschriebenen Clusterneubildung. Dabei müssen zwei allgemeine Fälle unterschieden werden: Gilt für die Subcluster C_{i_1} und C_{i_2} des Clusters C_i

$$n^{\alpha^{\mathrm{A}}}_{i.} \geq \beta^{\min}_{CT} n^{\alpha^{\mathrm{A}}}_{\min} \quad \wedge PD_{i.} \geq \beta^{\min}_{CT} PD_{\min}, \tag{5.46}$$

$\beta^{\min}_{CT} \in [0,1]$, ist das Cluster zumindest als potentiell teilbar vorzumerken. Gilt außerdem

$$n^{\alpha^{\mathrm{A}}}_{i.} \geq \beta_{CT} n^{\alpha^{\mathrm{A}}}_{\min} \quad \wedge PD_{i.} \geq \beta_{CT} PD_{\min}, \tag{5.47}$$

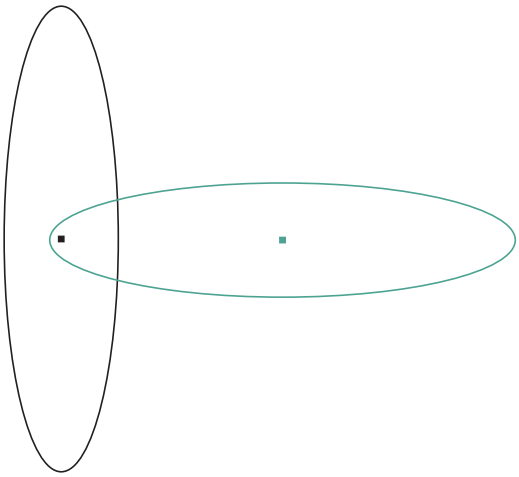


Abbildung 5.20.: Einbeziehung der Clusterausrichtung

$\beta_{CT} \in \left[\beta_{CT}^{\min}, 1\right]$, ist die Trennung als vollzogen anzunehmen.

Das Vorgehen für die mögliche Trennung eines Clusters lässt sich wie folgt zusammenfassen:

1. *Überprüfung relevanter Kriterien*:
   Gilt ein Cluster gemäß Abschnitt 5.5 aufgrund eines Dichterückgangs als gefährdet (vgl. (5.27)), erfolgt eine Prüfung der Änderungen bzgl. Kompaktheit und Clusterausrichtung bzw. -ausdehnung. Deuten die Veränderungen nicht auf eine mögliche Clusterteilung hin, wird das Cluster weiterhin als gefährdet betrachtet und im Rahmen der Clusterelimination näher untersucht.

2. *Subclustering*:
   Sind die relevanten Kriterien für eine potentielle Clusterteilung erfüllt, erfolgt eine Clusteranalyse unter Einbeziehung der vom aktuellen Cluster mit einem Zugehörigkeitsgrad von α_{SC}^{A} absorbierten Objekte für zwei Cluster. Liegt keine ausreichende Separierung vor, d.h., ihr Überschneidungsgrad überschreitet einen vorgegebenen Grenzwert $\lambda_{CT}^{\mathrm{I}^{\max}}$, ist keine Clustertrennung möglich; das Cluster wird weiterhin als gefährdet angesehen (s.o.).

3. *Prüfung einer ausreichenden Clustertrennung*:
   Werden zwei anhand ihrer auf dem Inklusionsmaß basierenden Ähnlichkeit unterscheidbare Cluster aufgedeckt, erfolgt eine Prüfung der Struktur der einzelnen Subcluster bzgl. der Anzahl absorbierter Objekte und ihrer Dichte sowie eine Evaluation der Separation der Clusterzentren. Nur bei Erfüllung aller Kriterien gilt das Cluster als bereits geteilt, ansonsten ist es bei ausreichender Mindestgröße und -dichte als potentiell teilbar zu markieren und gilt nicht länger als gefährdet.

5.7.2. Untersuchung der Entwicklung vorgemerkter Cluster

Wie bei den zuvor beschriebenen abrupten Veränderungen ist auch im Zusammenhang möglicher Clusterteilungen die zeitliche Entwicklung für eine Trennung vorgemerkter Cluster elementar, um den Zeitpunkt der Clustertrennung prognostizieren zu können. Wie in Abschnitt 5.7.1

beschrieben, werden Cluster dann für eine potentielle Teilung vorgemerkt, wenn eine ausreichende Clusterseparierung mittels eines Subclusterings möglich ist, jedoch entweder Objektzahl oder Dichte der Subcluster nicht ausreichen oder diese zu nah zueinander liegen, als dass sie als getrennt angesehen werden können. Zeigen die Subcluster eine positive Entwicklung im Kontext einer Clustertrennung, d.h., nähern sie sich im Zeitablauf nicht wieder an, muss der zu erwartende Trennungszeitpunkt $\hat{t}_i^{CT}$ prognostiziert werden. Solange die Subcluster ausreichend separiert sind, jedoch nicht genügend Objekte absorbiert werden oder ihre Dichte zu gering ist, ist unter Einbeziehung eines geeigneten Regressionsmodells der Zeitpunkt $\hat{t}_i^{CT}$ zu prognostizieren, von dem an eine ausreichende Erfüllung der Bedingung aus (5.47) angenommen wird.

Ist die Trennung zwischen den Subclustern hingegen noch nicht ausreichend, erfolgt eine separate Prüfung der Entwicklung der Subclustertrennung anhand der Clustertrajektorien der Subcluster und ihrer Überschneidungsgrade, sofern diese aus den vorherigen Analysezeitpunkten bekannt sind. Mit Hilfe eines geeigneten Regressionsmodells für die einzelnen Trajektorien ist es möglich, den Zeitpunkt $\hat{t}_i^{CT'}$ vorherzusagen, an dem eine ausreichende Trennung zwischen den Subclustern zu erwarten ist. Anschließend muss überprüft werden, ob zum Zeitpunkt $\hat{t}_i^{CT'}$ die Objektzahl und die Dichte der Subcluster ausreichen. Wird diese Bedingung erfüllt, so kann als Trennungszeitpunkt $\hat{t}_i^{CT} = \hat{t}_i^{CT'}$ angenommen werden; ansonsten ist analog zum zuvor beschrieben Vorgehen der Zeitpunkt $\hat{t}_i^{CT}$ zu schätzen, von dem an diese Bedingung voraussichtlich erfüllt sein wird.

Stagniert die Entwicklung bzgl. Objektzahl und Dichte der Subcluster bzw. ihrer Distanz oder ist sogar rückläufig, ist die Vormerkung zu verwerfen und das Cluster nicht länger als potentiell teilbar zu betrachten.

5.7.3. Veranschaulichung des Vorgehens anhand künstlicher Daten

Das konkrete Vorgehen zur Clustertrennung wird im Folgenden anhand eines vereinfachten Beispiels verdeutlicht. Analog zu den vorangegangenen Beispielen wurde ein zweidimensionaler Satz normalverteilter Daten kreiert, der aus drei eindeutig getrennten Clustern bestand. Dabei setzte sich eines der Cluster aus zwei nicht separierten, nahezu vollständig überlappenden Subclustern (Cluster 2 und Cluster 4) zusammen, an denen die Clustertrennung veranschaulicht werden soll. Den zwei deutlich getrennten Clustern 1 und 3 wurden jede Periode konstant 100 Objekte, den Subclustern des zusammengesetzten Clusters jeweils 75 neue Objekte hinzugefügt; die Werte für die Clusterzentren und Kovarianzmatrizen sind Tabelle 5.22 zu entnehmen. Das zweite und das vierte Cluster bilden zusammen das zu trennende Cluster; die Teilung wurde durch ein Bewegen des vierten Clusterzentrums um jeweils $\begin{pmatrix} -0.3 \\ 0 \end{pmatrix}$ je Periode umgesetzt. Zur Erhöhung der Clusterhomogenität im Ausgangscluster wurde für Cluster 2 gegenüber den vorherigen Beispielen eine Verringerung des Volumens vorgenommen. Insgesamt wurden zwölf Perioden betrachtet, wobei die Zeitfensterlänge entsprechend den vorangegangenen Beispielen bei $\tau = 3$ lag, die Analysefrequenz wurde auf $\Delta t = 1$ gesetzt.

Zur Analyse wurde jeweils der Gustafson-Kessel-Algorithmus für ellipsoide Cluster unter Einbeziehung der Dreiecksbeziehung der Distanzen mit $\alpha = 0.5$ angewandt (vgl. Abschnitt 4.3.3, Algorithmus 4.7); die η_i wurden einmalig neu berechnet. Neben dem benötigten Absorptions-

Cluster	Zentrum	Kovarianzmatrix
Cluster 1	$\begin{pmatrix} 2 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$
Cluster 2	$\begin{pmatrix} 8 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 1.2 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$
Cluster 3	$\begin{pmatrix} 12 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 0.8 & -0.5 \\ -0.5 & 0.6 \end{pmatrix}$
Cluster 4	$\begin{pmatrix} 8 \\ 13 \end{pmatrix}$	$\begin{pmatrix} 1.0 & -0.2 \\ -0.2 & 0.4 \end{pmatrix}$

Tabelle 5.22.: Vorgegebene Parameter

grenzwert $\alpha^{\mathrm{A}} = 0.125$ wurde zur Differenzierung der Teilung von einer möglichen Clusterelimination anhand des Kompaktheitsindexes $\alpha_{\kappa}^{\mathrm{A}} = 0.25$ gewählt; die Werte zur Evaluation der Änderungen bzgl. Ausrichtung und Ausdehnung wurden auf $\lambda_{par}^{CT} = 0.99$ bzw. $\lambda_{\theta}^{CT} = 0.01$ gesetzt. Das Subclustering erfolgte unter Einbeziehung aller Objekte zum potentiell teilbaren Cluster, die einen Zugehörigkeitsgrad von mindestens $\alpha_{SC}^{\mathrm{A}} = 0.08$ zu diesem Cluster besaßen. Die Prüfung einer ausreichenden Clusterseparation erfolgte unter Verwendung einer absoluten, kontextabhängig gewählten Mindestdistanz von $\lambda_{dist}^{CT} = 4.0$. Als potentiell trennbar wurden Cluster mit einer maximalen Überlappung von $\lambda_{CT}^{\mathrm{I}^{\max}} = 0.8$ betrachtet, bereits getrennte Cluster durften basierend auf dem Inklusionsmaß eine Ähnlichkeit von $\lambda_{CT}^{\mathrm{I}^{\min}} = 0.4$ nicht überschreiten. Zur Evaluation der Objektzahl und der Dichte der Subcluster im Vergleich zu den übrigen Clustern wurden $\beta_{CT}^{\min} = 0.3$ und $\beta_{CT} = 0.5$ gewählt.

Aufgrund der starken Überlappung bei nahezu deckungsgleichen Clusterzentren der Cluster 2 und 4 ist es in den ersten Perioden nicht möglich, Veränderungen innerhalb der Struktur des gemeinsamen Clusters aufzudecken; die einzelnen Subcluster stellen weiterhin ein gemeinsames Cluster dar. Erst nach einigen Perioden, wenn sich die Überschneidung der Subcluster deutlich verringert hat und so eine einheitliche Entwicklung einzelner Objekte deutlich wird, ist es möglich, die Veränderung innerhalb dieses Clusters aufzudecken. Aus diesem Grund erfolgt die Darstellung der Ergebnisse erst ab Periode $t = 8$.[24]

Abbildung 5.21 zeigt die Analyseergebnisse. Zum Zeitpunkt $t = 8$ (Abbildung 5.21a) werden wie erwartet drei gut separierte Cluster erkannt; die Darstellung des oberen Clusters erfolgt dabei über eine farbliche Trennung der Objekte der Subcluster. In Abbildung 5.21b kann bereits zum Folgezeitpunkt $t = 9$ eine leichte Ausdehnung des zusammengesetzten Clusters bei gleichzeitig zurückgehender Dichte im Clusterzentrum verzeichnet werden, während die separierten Cluster nahezu unverändert sind. Dies zeigt sich anhand eines deutlichen Rückgangs der Fuzzy-Kardinalität um 11.5282% bei gleichzeitiger Zunahme des Fuzzyhypervolumens um 9.4053%. Unter Hinzunahme des Kompaktheitsindexes nach Bensaid u. a. (1996) wird die Umverteilung der Objekte innerhalb des Clusters deutlich: Für $\alpha^{\mathrm{A}} = 0.125$ nimmt der Kompakt-

[24] Die langsam verlaufende Clusterteilung lässt sich nur bedingt beschleunigen. Würde das Auseinanderdriften des Clusters schneller stattfinden, indem sich Cluster 4 in größeren Schritten von Cluster 2 entfernte, ginge dies unter Umständen auf Kosten der Nachvollziehbarkeit der Entwicklung innerhalb des Clusters. Da die Objekte des vierten Clusters den durch α^{A} bzw. $\alpha_{\kappa}^{\mathrm{A}}$ begrenzten Betrachtungsraums nahezu abrupt verließen, würde eher ein Rückgang der allgemeinen Objektzahl im gemeinsamen Cluster verzeichnet, während an anderer Stelle ein neu entstandenes Cluster aufgedeckt würde.

heitsindex gegenüber den Analyseergebnissen aus $t = 8$ um 6.0164% zu, der entsprechende Wert für $\alpha_\kappa^{\mathrm{A}} = 0.25$ steigt um 6.0623% an. Da sich die Teilung ungefähr entlang der Ausrichtung des Clusters vollzieht, kann keine sichtbare Drehung verzeichnet werden; die Anpassung der Clusterausdehnung wird jedoch anhand der Eigenwerte deutlich, deren Verhältnis sich um 0.0123 (entsprechend 6.3174%) ändert. Eine Teilung liegt also im Bereich des Möglichen. Das Subclustering zeigt jedoch, dass keine ausreichende Separation vorliegt, um die Subcluster als eigenständige Cluster anzusehen; die durch das Subclustering ermittelten Zentren sind in Tabelle 5.23 dokumentiert.

Cluster	**Zentrum**	**Distanz**
Cluster 2	$\begin{pmatrix} 6.3721 \\ 13.1530 \end{pmatrix}$	$d^2(\vec{v}_2, \vec{v}_4) = 0.6137$
Cluster 4	$\begin{pmatrix} 6.0546 \\ 12.4369 \end{pmatrix}$	

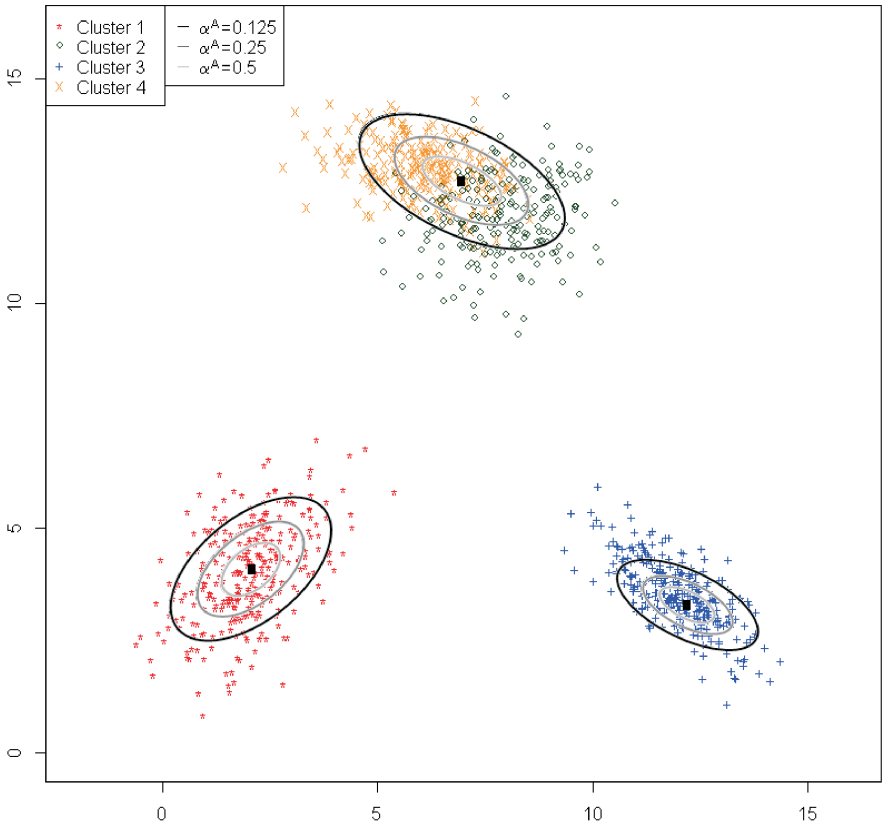
Tabelle 5.23.: Subclusterzentren zum Zeitpunkt $t = 9$

Zum darauffolgenden Zeitpunkt $t = 10$ stellt sich die Entwicklung des zusammengesetzten Clusters in Abbildung 5.21c aufgrund der Verschiebung eines der Subcluster und der Umverteilung der Objekte nach außen optisch noch deutlicher dar, da sich eine eindeutige Dehnung erkennen lässt. Dies führt zu einer Reduzierung der Fuzzy-Kardinalität um 6.6708%; ferner nimmt das Fuzzyhypervolumen gegenüber der vorangegangen Analyse um 7.3336% zu. Auch die Kompaktheitsindizes demonstrieren die stattgefundene Entwicklung: Bei einem größeren Betrachtungsraum mit $\alpha^{\mathrm{A}} = 0.125$ wächst der Index um 2.0108%, bei kleinerem Betrachtungsraum und der Fokussierung auf das Clusterzentrum mit $\alpha_\kappa^{\mathrm{A}} = 0.25$ steigt er um 4.5767%. Die durch die Verschiebung des Subclusters hervorgerufene Ausdehnung des Clusters manifestiert sich in einer Veränderung des Eigenwertverhältnisses von 0.0226, welches einer Anpassung um 23.0143% entspricht; da die Änderung entlang der Clusterausrichtung stattfindet, kann abermals keine sichtbare Drehung des Clusters festgestellt werden. Entsprechend dem vorigen Analysezeitpunkt gilt das Cluster als potentiell trennbar, die Separation der Subcluster reicht jedoch nach wie vor nicht aus, eine tatsächliche Clusterteilung durchzuführen. Die durch das Subclustering resultierenden Werte für die Clusterzentren sind in Tabelle 5.24 angegeben.

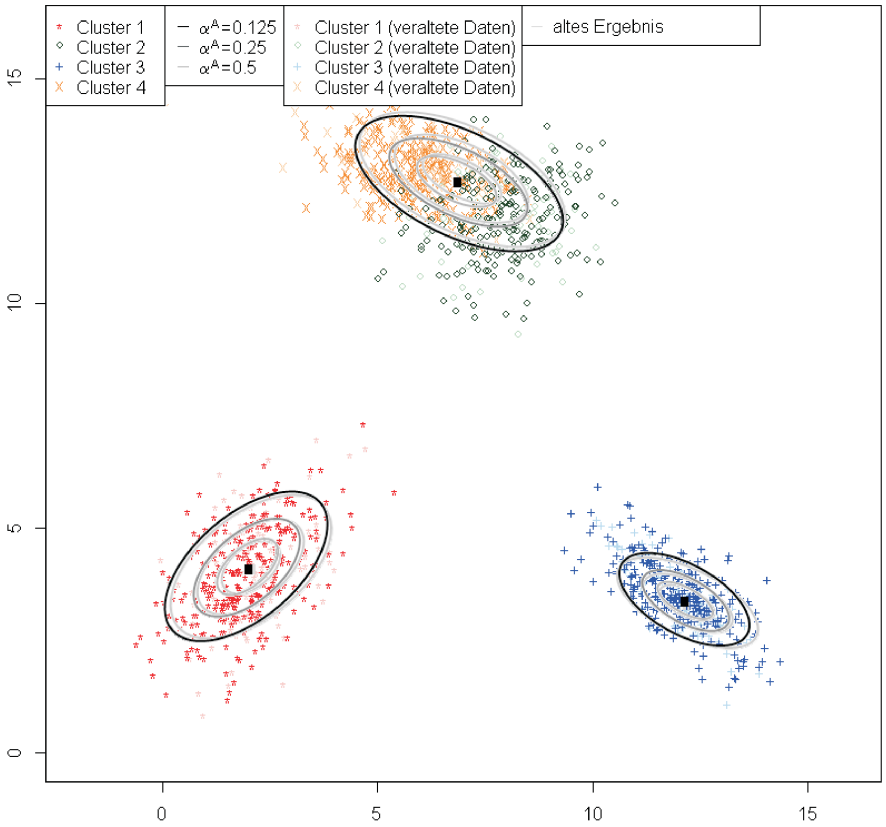
Cluster	**Zentrum**	**Distanz**
Cluster 2	$\begin{pmatrix} 6.5513 \\ 13.0084 \end{pmatrix}$	$d^2(\vec{v}_2, \vec{v}_4) = 1.1521$
Cluster 4	$\begin{pmatrix} 5.5517 \\ 12.6174 \end{pmatrix}$	

Tabelle 5.24.: Subclusterzentren zum Zeitpunkt $t = 10$

Da nun Werte aus zwei Perioden bzgl. der Entwicklung der Subcluster vorhanden und die Bedingungen bzgl. Überschneidungsgrad der Subcluster, Anzahl absorbierter Objekte und deren Dichte je Subcluster bereits zuvor erfüllt worden sind, ist zu diesem Zeitpunkt eine erste Prognose zwar möglich, aber nur bedingt sinnvoll. Unterstellt man eine konstante, lineare Entwicklung der Clusterzentren voneinander fort, ergäbe sich erst für den Zeitpunkt $t = 12$ eine ausreichende Distanz zwischen den Subclusterzentren, dargestellt in Tabelle 5.25.

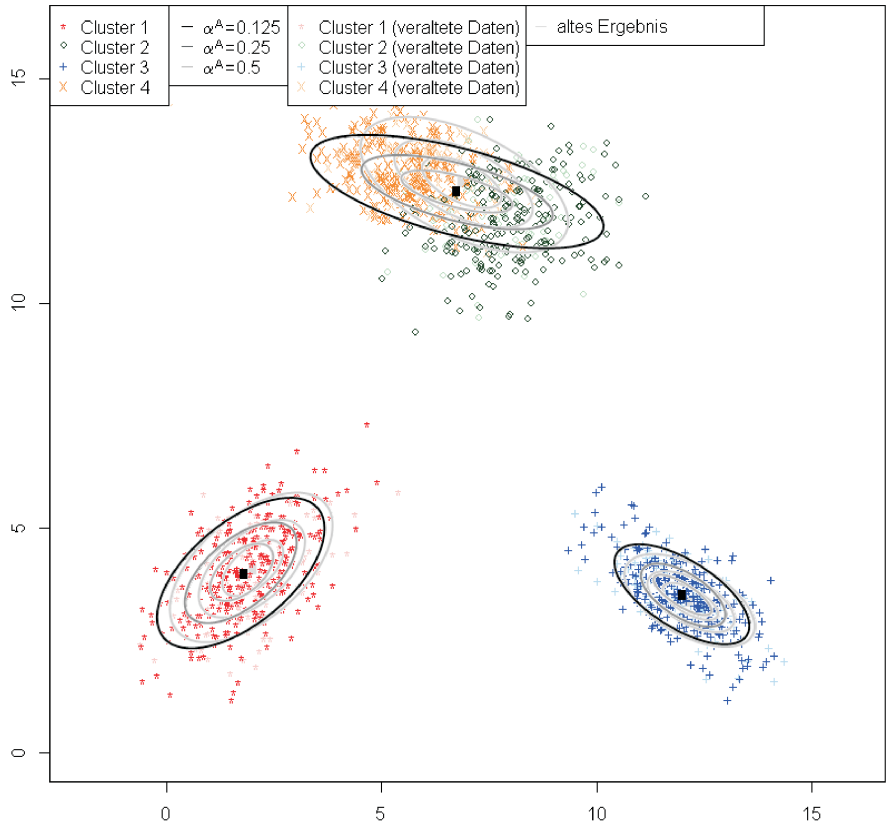


(a) Clusterstruktur für den Zeitpunkt $t = 8$

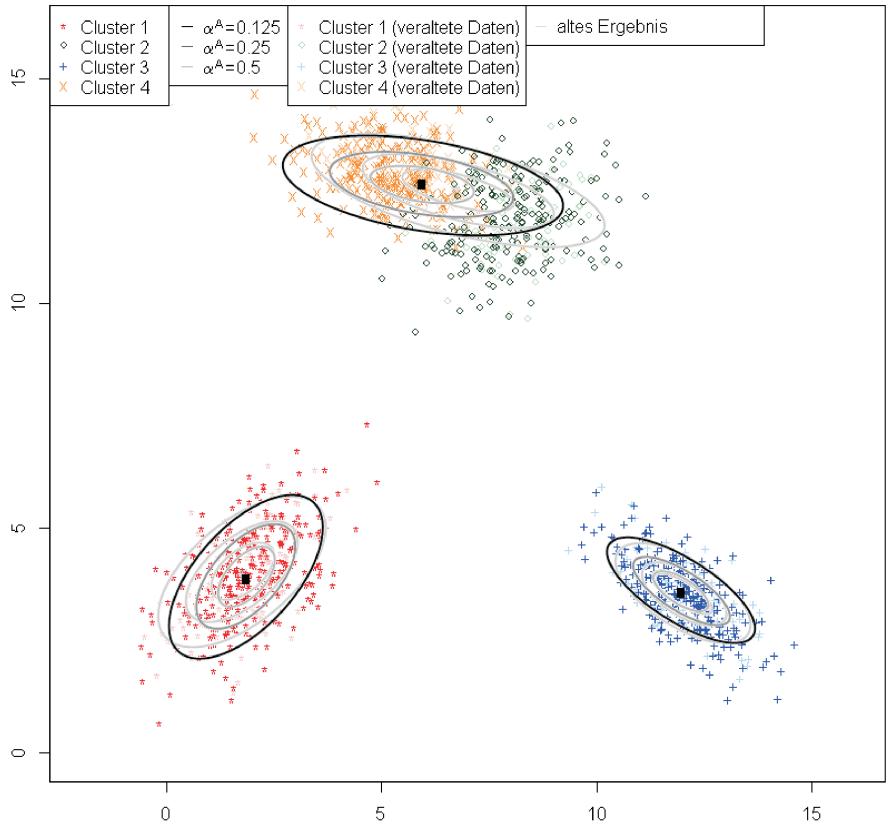


(b) Clusterstruktur für den Zeitpunkt $t = 9$

Abbildung 5.21.: Beispiel für Clustertrennung

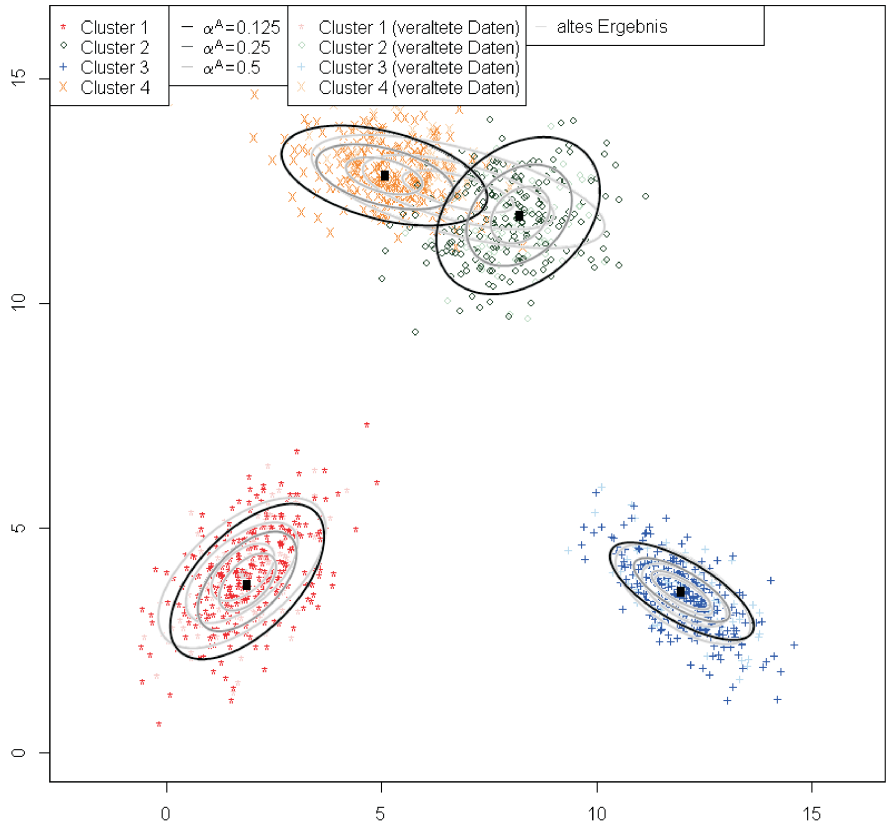


(c) Clusterstruktur für den Zeitpunkt $t = 10$

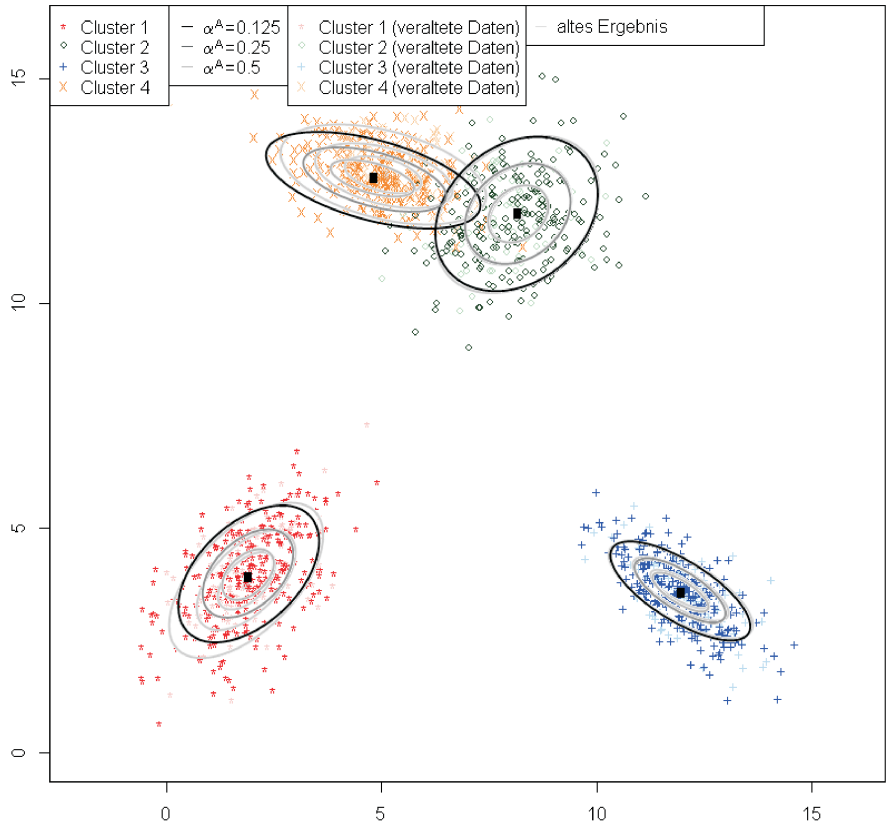


(d) Clusterstruktur für den Zeitpunkt $t = 11$

Abbildung 5.21.: Beispiel für Clustertrennung



(e) Clusterstruktur für den Zeitpunkt $t = 11$ (nach Reclustering)



(f) Clusterstruktur für den Zeitpunkt $t = 12$

Abbildung 5.21.: Beispiel für Clustertrennung

	Cluster	Geschätztes Zentrum	Distanz
$t = 11$	Cluster 2	$\begin{pmatrix} 6.7305 \\ 12.8638 \end{pmatrix}$	$d^2\left(\hat{\vec{v}}_2, \hat{\vec{v}}_4\right) = 2.8325$
	Cluster 4	$\begin{pmatrix} 5.0488 \\ 12.7979 \end{pmatrix}$	
$t = 12$	Cluster 2	$\begin{pmatrix} 6.9097 \\ 12.7192 \end{pmatrix}$	$d^2\left(\hat{\vec{v}}_2, \hat{\vec{v}}_4\right) = 5.6547$
	Cluster 4	$\begin{pmatrix} 4.5459 \\ 12.9784 \end{pmatrix}$	

Tabelle 5.25.: Prognostizierte Subclusterzentren

Die Unterstellung einer solch linearen Entwicklung ist jedoch nicht geeignet: Aufgrund der wachsenden Distanz zwischen den Subclusterzentren nimmt der Einfluss der Objekte des jeweils anderen Subclusters bei der Bestimmung des Clusterzentrums ab, so dass eine schnellere Trennung zu erwarten ist. Da aber nur Werte für zwei Perioden vorliegen, ist eine geeignete Modellwahl und damit eine genauere Vorhersage bisher nicht möglich.

Die Überschätzung der Zeitspanne bei Anwendung eines linearen Modells zur Bestimmung des Trennungszeitpunkts zeigt sich bereits zum Zeitpunkt $t = 11$ (Abbildung 5.21d): Aufgrund der weiteren Ausdünnung verschiebt sich das gemeinsame Clusterzentrum in Richtung des dichteren Subclusters. Durch diese Entwicklung sind nur noch geringe Änderungen der Kardinalitäten zu verzeichnen, während das Fuzzyhypervolumen weiterhin steigt (7.6714%). Eine Differenzierung zur Clusterelimination ist an dieser Stelle mit Änderungen von $\kappa^{0.125}$ um 1.2647% bzw. $\kappa^{0.25}$ um 1.5246% nur noch sehr eingeschränkt möglich, da die Anpassung unter verstärkter Berücksichtigung des dichteren Subclusters auch hier kaum deutlich wird. Weil das Cluster jedoch als potentiell trennbar bekannt ist, ist eine Differenzierung nicht länger erforderlich. Anhand der Ausdehnung lässt sich die vorhandene Änderung abbilden: Das Verhältnis der Eigenwerte ändert sich um 0.0222, d.h. um 19.6983%; eine sichtbare Drehung findet wie zuvor nicht statt. Das Subclustering ergibt, dass die Teilung bereits vollzogen wurde, die Separation zwischen den Subclustern dementsprechend ausreicht. Die konkreten Werte sind in Tabelle 5.26 dokumentiert.

Cluster	Zentrum	Distanz
Cluster 2	$\begin{pmatrix} 7.8492 \\ 12.0928 \end{pmatrix}$	$d^2\left(\vec{v}_2, \vec{v}_4\right) = 8.1955$
Cluster 4	$\begin{pmatrix} 5.0964 \\ 12.8787 \end{pmatrix}$	

Tabelle 5.26.: Subclusterzentren zum Zeitpunkt $t = 11$

Die Ergebnisse zeigen deutlich den abnehmenden Einfluss des dichteren Subclusters auf das Subcluster geringerer Dichte: Das dichtere Subcluster (Cluster 4) entwickelt sich nahezu linear weiter, während Cluster 2 eine sprunghafte Entwicklung in Richtung des intuitiven Clusterzentrums vollführt. Das Ergebnis des resultierenden Reclusterings ist in Abbildung 5.21e dargestellt. Zum Zeitpunkt $t = 12$ in Abbildung 5.21f kann eine weitere leichte Verschiebung von

Cluster 4 festgestellt werden, die übrigen Cluster – inklusive des nun getrennten Cluster 2 – sind unverändert. Die Analyse der graduellen Unterschiede ergibt, dass die Trennung angemessen war, da keine mögliche Clustervereinigung angezeigt wird.

5.8. Zusammenfassung der Analyseschritte

Zur besseren Übersicht wird im Folgenden eine kurze Darstellung der einzelnen Analyseschritte vorgenommen, die in Abbildung 5.22 anhand eines Ablaufplans nachzuvollziehen sind. Dabei ist zu beachten, dass es sich zum besseren Verständnis um eine stark vereinfachte Darstellung der Untersuchung handelt: Die einzelnen Updates und Prüfungen bzgl. abrupter Veränderungen müssen für jedes Cluster separat erfolgen, bevor eine neue Clusterzahl bestimmt und ein eventuelles Reclustering durchgeführt werden kann. Außerdem ist neben der aufgeführten Prüfung vorhandener abrupter Unterschiede auch ihre Prognose gemäß den vorherigen Abschnitten 5.4 bis 5.7 durchzuführen.

Bevor mit der eigentlichen Untersuchung aufgetretender Unterschiede begonnen werden kann, werden die in den letzten Δt Perioden neu hinzugekommen Objekte anhand der aus der vorherigen Analyse bekannten Clusterstruktur den einzelnen Clustern zugeordnet.

Da graduelle Änderungen aktuell relevanter Objekte bereits spätere abrupte Veränderungen implizieren können, erfolgt ihre Untersuchung vor der Evaluation abrupter Anpassungen. Durch dieses Vorgehen kann auch, wie in Abschnitt 5.3 erläutert, einem irrtümlichen Aufdecken abrupter Änderungen entgegengewirkt werden. Bei der Analyse gradueller Unterschiede werden zunächst die inkrementellen Updates der einzelnen Clusterprototypen vorgenommen. Erweisen sich die auftretenden Verschiebungen innerhalb eines Clusters als signifikant, wird der geschätzte Clusterprototyp anstelle des bisher bekannten für die folgenden Analyseschritte übernommen; andernfalls wird der ursprüngliche Prototyp beibehalten.

Auf Basis der aktualisierten Clusterstruktur erfolgt die Untersuchung der graduellen Änderungen innerhalb der einzelnen Cluster, um Unterschiede bzgl. der Kardinalitäten, der Dichte, des Volumens und der Ausdehnung aufdecken zu können. Werden hier entsprechend Abschnitt 5.3.2 signifikante Veränderungen aufgedeckt, müssen die durch die lokalen Strukturmaße implizierten abrupten Veränderungen genauer überprüft werden (vgl. Tabelle 5.8). Liegt eine Gefährdung eines Clusters vor, muss zudem die als Verursacher der Clustergefährdung in Frage kommende abrupte Veränderung verifiziert werden (vgl. Abschnitte 5.5 und 5.7), bevor mit der Analyse fortgefahren werden kann. Die Prüfung neu entstehender Cluster erfolgt unabhängig von dem Vorhandensein signifikanter gradueller Änderungen, da im Rahmen der graduellen Unterschiede lediglich in der Clusterstruktur vorkommende Cluster untersucht und keine Ausreißer beachtet werden.

Werden abrupte Änderungen aufgedeckt, werden die Clusterzahl aktualisiert sowie ein Reclustering durchgeführt. Sind keine abrupten Veränderungen zu berücksichtigen, jedoch signifikante graduelle Änderungen bzgl. der Clusterprototypen vorhanden, so kann auch in diesem Fall ein Reclustering von Nutzen sein, um die aktuelle Positionierung und Ausrichtung der Cluster zu spezifizieren. Das Reclustering entfällt, wenn keine relevanten Unterschiede im Vergleich zur vorherigen Analyse gefunden werden; die alte Clusterstruktur bleibt in diesem Fall

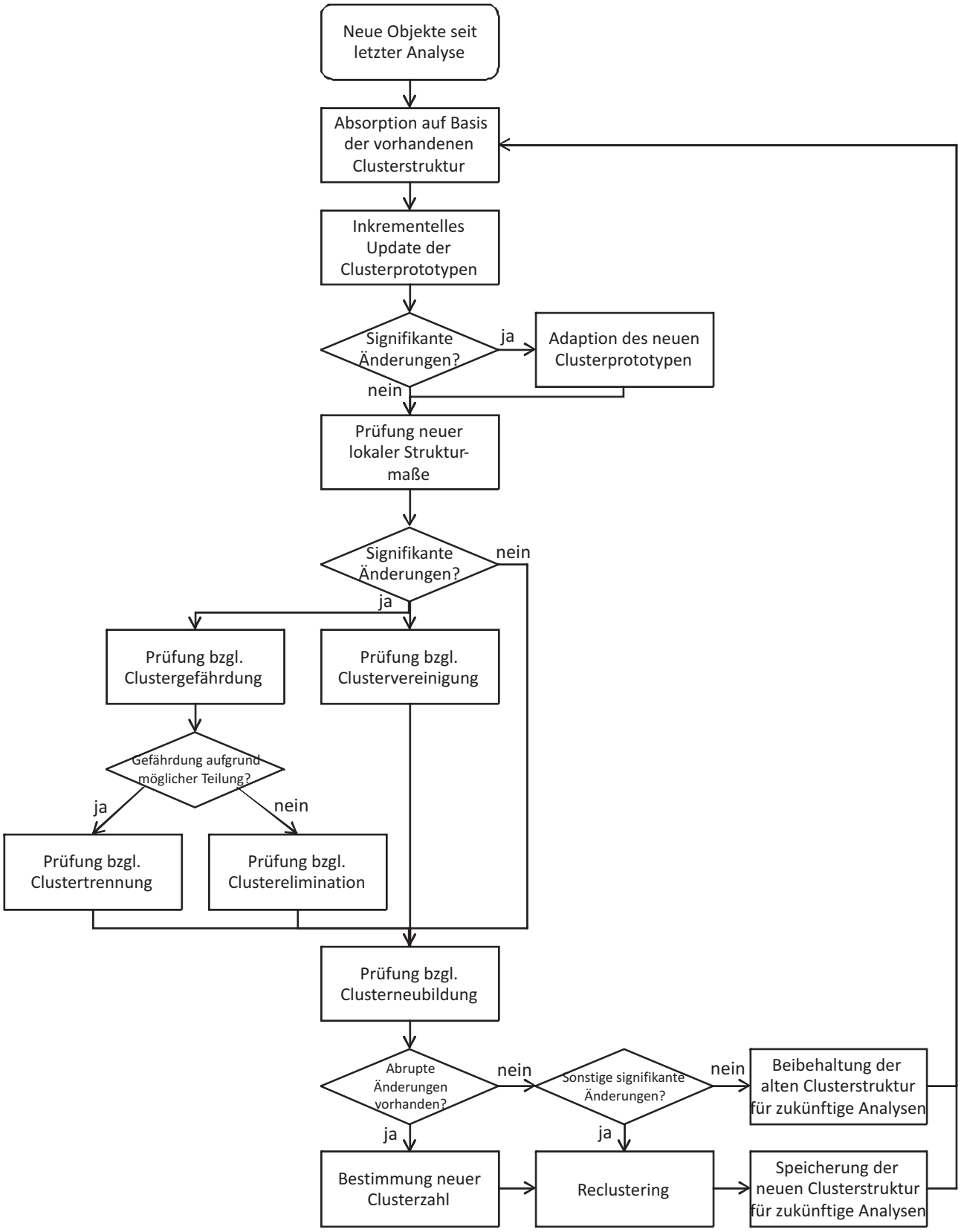


Abbildung 5.22.: Ablaufplan der Analyseschritte

für zukünftige Analysen erhalten.

Werden für die möglichen abrupten Anpassungen nur potentiell künftige Veränderungen herausgestellt, erfolgt ihre Analyse im Rahmen der in den einzelnen Abschnitten beschriebenen Prüfungen der jeweiligen Änderung.

Das geschilderte Vorgehen bietet natürlich keine Sicherheit, dass die Vorhersage zukünftig eintretender Anpassungen in jedem Fall exakte Ergebnisse liefert. Insbesondere bei plötzlich auftretenden Änderungen gelingt eine prediktive Vorgehensweise kaum; auch bei sehr schleichend stattfindenen Änderungen innerhalb der Clusterstruktur wird eine Prognose abhängig von den gewählten Parametergrenzen bei der graduellen Änderungen schwierig. Außerdem sind Wechselwirkungen einzelner abrupter Unterschiede denkbar. So besteht z.B. die Möglichkeit, dass die Teilung eines Clusters nicht aufgedeckt, stattdessen aber die Wanderung eines Clusterteils bei gleichzeitiger Neuentstehung eines zweiten Clusters sichtbar wird. Daher ist neben der automatischen Analyse der Änderungen die kritische Betrachtung der Analyseergebnisse elementar.

Wenn du keine Fehler machst, dann sind die Probleme, an denen du arbeitest, nicht schwierig genug. Und das ist ein großer Fehler.

F. Wilczek

6 Zusammenfassung und Ausblick

Das Aufdecken von Strukturen in Datensätzen ist in verschiedenen Bereichen von grundlegender Bedeutung. In der Medizin kann sich die zielgerichtete Datenanalyse z.B. bei der Diagnose bestimmter Krankheiten als hilfreich erweisen, im technischen Bereich unterstützt sie das Aufdecken einzelner Fehlerquellen. Im Marketingkontext und insbesondere in der Marktforschung ist das Data Mining elementar, um die Bedürfnisse am Markt aufdecken und geeignete Aktionen einleiten zu können. Jedoch erweist sich eine rein deskriptive Auswertung der aktuellen Situation nicht in jedem Fall als ausreichend, vielmehr sollten zukünftig zu erwartende Entwicklungen in die aktuelle Planung einbezogen werden. Dabei gilt es, auf Basis vorhandener Vergangenheitsdaten im Sinne des Change Minings existierende Unterschiede aufzudecken und daraus resultierende zukünftige Veränderungen und Ereignisse zu prognostizieren, um mögliche Reaktionen ableiten zu können. Die Untersuchung von Änderungen in der Datenstruktur kann je nach Untersuchungsgegenstand und Zielsetzung anhand verschiedener Data Mining-Verfahren erfolgen, so z.B. durch die im Marketing verbreiteten Methoden zur Multidimensionalen Skalierung und zur Ermittlung von Assoziationsregeln, die zu Beginn dieser Ausführungen exemplarisch beschrieben wurden.

Deutlich komplexer wird die Aufgabenstellung jedoch, wenn Veränderungen innerhalb einer Gruppierung ähnlicher Objekte aufgrund ihrer Eigenschaften wie z.B. das Erkennen ähnlichen Kundenverhaltens in einzelnen Segmenten nicht nur aufgedeckt, sondern auch für die Zukunft vorhergesagt werden sollen. Die Auswahl spezieller Methoden, um ein Monitoring der vorhandenen Unterschiede zwischen den verschiedenen Untersuchungszeitpunkten durchführen und darauf aufbauend spätere Entwicklungen prognostizieren zu können, ist unabdingbar. Für die Einteilung ähnlicher Datensätze in homogene Gruppen wird die Clusteranalyse angewandt, in deren Rahmen die Ähnlichkeit der Eigenschaftsausprägungen einzelner Objekte analysiert wird. Gerade bei der Analyse dynamischer Änderungen erweist es sich als elementar, auch graduelle Unterschiede zwischen den einzelnen Untersuchungszeitpunkten zu ermitteln; aus diesem Grund erscheint eine Anwendung harter Clusterverfahren, die ein Objekt eindeutig genau einem Cluster zuordnen, als nicht zielführend. Um auch langsam verlaufende Änderungen verdeutlichen zu können, werden vielmehr graduelle Abstufungen in den Zugehörigkeiten benötigt, sogenannte Zugehörigkeitsgrade. Dabei empfehlen sich insbesondere die Ansätze zur possibilistischen Fuzzy-Clusteranalyse, da so die Datenstruktur aufgrund der Beurteilung der Typizität einzelner Objekte für die unterschiedlichen Cluster exakter abgebildet werden kann. Dementsprechend wird damit sowohl die Identifikation von Ausreißern ermöglicht wie auch das Hervorheben derjenigen Objekte, die gleichermaßen mehreren Clustern zuzuordnen sind. Um die irrtümliche

Darstellung sich teilweise überlappender Cluster als gemeinsames Cluster zu vermeiden, erweist sich bei der ursprünglichen possibilistischen Fuzzy-Clusteranalyse jedoch die notwendige deutlich erkennbare Trennung als problematisch. Eine solch eindeutig vorhandene Separation liegt in verschiedenen praktischen Anwendungsbereichen i.d.R. nicht vor, weil diverse Objekte keine klare Zugehörigkeit zu den einzelnen Gruppen vorweisen können. In der Literatur finden sich daher unterschiedliche Ansätze zur Erweiterung zur possibilistischen Fuzzy-Clusteranalyse. Diese Ansätze fokussieren jedoch zumeist auf die possibilistische Analyse statischer Strukturen und sind deswegen für die Analyse im Rahmen des Change Minings nur bedingt geeignet. Der eingeführte Ansatz von Timm u. a. (2001) basierend auf der Abstoßung einzelner Clusterzentren verhindert beispielsweise, dass Cluster sich über die Zeit hinweg annähern und schließlich vereinigt werden können. Um diesem Problem entgegenzuwirken, wurde ein neuer Ansatz zur Erweiterung des Originalansatzes zur possibilistischen Clusteranalyse vorgestellt, der die Homogenität innerhalb einzelner Cluster einbezieht, indem er dichtere Regionen hervorhebt und Regionen von geringerer Dichte innerhalb eines Clusters in der Zielfunktion bestraft. Auf diese Weise ist es möglich, auch im dynamischen Kontext Änderungen in der Clusterstruktur aufzudecken.

Mit Hilfe des vorgestellten Ansatzes zur possibilistischen Clusteranalyse kann eine Analyse einer Clusterstruktur im Rahmen des Change Minings so durchgeführt werden, dass die Möglichkeit zur Vorhersage zukünftiger Änderungen innerhalb einer Clusterstruktur besteht. Dabei liegt der Fokus auf der Untersuchung dynamischer Clusterstrukturen, wobei die Objekte durch zum Untersuchungszeitpunkt statische, sich zwischen den einzelnen Zeitpunkten jedoch verändernde Eigenschaftsvektoren beschrieben sind. So können auch anonyme Objekte wie unbekannte Kunden eines Geschäfts oder vertrauliche, anonymisierte Patientendaten untersucht werden, da die Veränderungen der einzelnen Objekte nicht bedeutungsrelevant sind.

Es gibt verschiedene Änderungen innerhalb einer Clusterstruktur, die beschrieben und prognostiziert werden müssen. Graduelle Unterschiede zwischen den einzelnen Untersuchungszeitpunkten können die Entwicklung bzgl. Position und clusterinterner Struktur einzelner Cluster verdeutlichen, während abrupte Veränderungen die generelle Clusterstruktur und die Anzahl der darin enthaltenen Cluster betreffen. Dabei können die graduellen Anpassungen auf zukünftig zu erwartende abrupte Änderungen hinweisen: Nimmt beispielsweise die Anzahl durch ein Cluster absorbierter Objekte stetig ab, steht möglicherweise eine Clusterelimination bevor. Bewegen sich hingegen einzelne Cluster aufeinander zu, d.h., die zugehörigen Objekte werden einander ähnlicher, kann diese Entwicklung als Indikator einer zukünftigen Clustervereinigung gelten. Daher sind graduelle Unterschiede vor abrupten zu untersuchen, damit die resultierenden Ergebnisse zur Untersuchung der abrupten Veränderungen hinzugezogen werden können. Zur Verdeutlichung gradueller Entwicklungen wurden verschiedene bekannte Maße zur Beschreibung der clusterinternen Struktur aus der Literatur übernommen und ggf. auf die speziellen Ansprüche im Change Mining angepasst. Ferner wurde eine Möglichkeit zum inkrementellen Update der Clusterprototypen vorgestellt und erweitert, damit auch die Entwicklung bzgl. der Clusterpositionierung nachvollzogen werden kann. Im Anschluss erfolgte eine Zuordnung der einzelnen graduellen Änderungen zu den durch sie implizierten abrupten Veränderungen.

Die Untersuchung der abrupten Anpassungen erfolgt einzeln für die verschiedenen Änderungstypen. Die Neubildung einzelner Cluster wird unabhängig von graduellen Entwicklungen innerhalb der allgemeinen Clusterstruktur untersucht, da in der vorhandenen Clusterstruktur

Ausreißer zwar keine explizite Berücksichtigung finden, gleichwohl zum Aufdecken sich in der Entstehung befindender Cluster elementar sind. Sobald ein potentiell entstehendes, jedoch noch nicht eigenständiges Cluster erkannt wird, muss dieses auch bei einem möglichen Reclustering der vorhandenen Objekte einbezogen werden, um einen Einfluss der das wachsende Cluster repräsentierenden Objekte auf die Lage der übrigen Cluster zu verhindern. Basierend auf den vorhandenen Informationen zu einzelnen Ausreißerclustern, die die Entstehung neuer Cluster verdeutlichen, gilt es, den Zeitpunkt vorherzusagen, zu dem ein Cluster als eigenständiges Cluster in der Clusterstruktur angesehen werden kann. Dieser Zeitpunkt kann mit Hilfe eines Regressionsmodells bestimmt werden, die dazu benötigten Grenzwerte sind jedoch kontextabhängig zu wählen.

Die übrigen Veränderungen zur Clusterelimination, Clustervereinigung und Clustertrennung basieren auf den Untersuchungen zu den graduellen Unterschieden einzelner Cluster. So können Dichte- und Volumenänderungen sowie eine Verringerung der Anzahl der durch ein Cluster absorbierten Objekte auf eine bevorstehende Clusterelimination hindeuten, ebenso können sie jedoch auch als Indikator für eine bevorstehende Clustertrennung dienen. Anhand der Kompaktheit und der Dichteanpassungen innerhalb des Clusters muss zwischen den einzelnen Änderungsarten differenziert werden, bevor die konkrete Entwicklung näher untersucht und der Zeitpunkt ihres Eintreffens unter Einbeziehung eines geeigneten Regressionsmodells prognostiziert werden kann. Hierzu wurden verschiedene Maße und Grenzwerte eingeführt, die die Unterscheidung der verschiedenen Typen unterstützen und im prädiktiven Kontext zur Analyse zukünftig zu erwartender Änderungen hinzugezogen werden können. Dabei sind die Grenzwerte analog zur Clusterneubildung kontextabhängig zu wählen, da beispielsweise bei verschiedenen theoretisch fundierten Untersuchungsgegenständen eine frühzeitige Unterscheidung einzelner Cluster erforderlich ist, während in einer Marktstruktur erst eine deutlich sichtbare Trennung einzelner Segmente vorhanden sein sollte, ehe eine differenzierte Bearbeitung erfolgt. Dasselbe gilt für die Vereinigung von Clustern, die basierend auf graduellen Änderungen zur Clusterpositionierung und -ausrichtung untersucht wird. Zu diesem Zweck wurden ebenfalls entsprechende Maße eingeführt und nach ihrer Priorität geordnet, so dass der Fokus auf der Überlappung einzelner Cluster liegt, um diese zusammenzufassen, jedoch auch ihre Ähnlichkeit basierend auf der Lage zweier Cluster zueinander berücksichtigt wird.

Zu allen auftretenden Änderungen wurde eine Vielzahl an Experimenten basierend auf verschiedenen Ausgangsdaten bzgl. Clusterzahl, Objektzahl je Cluster, Dimensionierung und zeitlichen Parametern durchgeführt (vgl. Anhang C). Ferner unterschied sich die Deutlichkeit der Separation zwischen den vorhandenen Clustern und damit das Vorkommen von Ausreißern in den einzelnen Experimenten, so dass ein breites Spektrum möglicher Entwicklungen abgebildet wurde. Diese Experimente wurden unter Anwendung eines eigens dafür entwickelten Java-Applets durchgeführt, das in Anhang C.1 detailliert erläutert wird. Für die bessere Nachvollziehbarkeit der konkreten Vorgehensweisen zur Untersuchung der einzelnen Entwicklungen wurde ein anschauliches Beispiel ausgewählt, das in den einzelnen Abschnitten zur Evaluation der Unterschiede herangezogen wurde. Trotz der umfangreichen Experimente kann die Exaktheit der Ergebnisse jedoch nicht in jedem Fall garantiert werden. Zum einen ist es nicht möglich, plötzlich auftretende Veränderungen vorherzusagen, zum anderen sind alle Untersuchungen abhängig von der gewählten Parameterwahl.

Obwohl aus den durchgeführten Experimenten deutlich wird, dass das vorgestellte Monitoring

der Unterschiede und die darauf basierende Vorhersage abrupter Veränderungen für das Change Mining sehr vielversprechend sind, steht die Übertragung auf praxisbezogene Daten aus. Bei der Wahl der für ein solches Praxisbeispiel verwendeten Daten ist zu beachten, dass eine heterogene Gesamtstruktur benötigt wird, in der im Zeitablauf Änderungen auftreten.

A

Herleitungen zu den Ansätzen der Clusterabstoßung

A.1. Herleitungen für Modellierung mittels Clusterabstoßung – Erweiterung 1

$$J_{Het1}\left(X,U,C\right)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_j\right)+\sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}\left(1-u_{ij}\right)^{m}+\sum_{i=1}^{c}\gamma_i\sum_{\substack{i'=1\\i'\neq i}}^{c}\frac{1}{\zeta d^{2}\left(C_i,C_{i'}\right)}$$

$$\begin{aligned}
\frac{\partial J_{Het1}\left(X,U,C\right)}{\partial\vec{v}_i} &= \sum_{j=1}^{n}u_{ij}^{m}\frac{\partial}{\partial\vec{v}_i}\|\vec{x}_j-\vec{v}_i\|_{A_i}^{2}+\frac{\gamma_i}{\zeta}\sum_{\substack{i'=1\\i'\neq i}}^{c}\frac{\partial}{\partial\vec{v}_i}\frac{1}{\frac{1}{2}\left(\|\vec{v}_{i'}-\vec{v}_i\|_{A_i}^{2}+\|\vec{v}_{i'}-\vec{v}_i\|_{A_{i'}}^{2}\right)}\\
&\quad+\sum_{\substack{i'=1\\i'\neq i}}^{c}\frac{\gamma_{i'}}{\zeta}\frac{\partial}{\partial\vec{v}_i}\frac{1}{\frac{1}{2}\left(\|\vec{v}_i-\vec{v}_{i'}\|_{A_{i'}}^{2}+\|\vec{v}_i-\vec{v}_{i'}\|_{A_i}^{2}\right)}\\
&= \sum_{j=1}^{n}u_{ij}^{m}\lim_{t\to0}\frac{\|\vec{x}_j-(\vec{v}_i+t\vec{\xi})\|_{A_i}^{2}-\|\vec{x}_j-\vec{v}_i\|_{A_i}^{2}}{t}\\
&\quad+\sum_{\substack{i'=1\\i'\neq i}}^{c}\frac{\gamma_i}{\zeta}\lim_{t\to0}\left(\frac{1}{t}\left(\frac{1}{\frac{1}{2}\left(\|\vec{v}_{i'}-(\vec{v}_i+t\vec{\xi})\|_{A_i}^{2}+\|\vec{v}_{i'}-(\vec{v}_i+t\vec{\xi})\|_{A_{i'}}^{2}\right)}\right.\right.\\
&\qquad\left.\left.-\frac{1}{\frac{1}{2}\left(\|\vec{v}_{i'}-\vec{v}_i\|_{A_i}^{2}+\|\vec{v}_{i'}-\vec{v}_i\|_{A_{i'}}^{2}\right)}\right)\right)\\
&\quad+\sum_{\substack{i'=1\\i'\neq i}}^{c}\frac{\gamma_{i'}}{\zeta}\lim_{t\to0}\left(\frac{1}{t}\left(\frac{1}{\frac{1}{2}\left(\|(\vec{v}_i+t\vec{\xi})-\vec{v}_{i'}\|_{A_{i'}}^{2}+\|(\vec{v}_i+t\vec{\xi})-\vec{v}_{i'}\|_{A_i}^{2}\right)}\right.\right.\\
&\qquad\left.\left.-\frac{1}{\frac{1}{2}\left(\|\vec{v}_i-\vec{v}_{i'}\|_{A_{i'}}^{2}+\|\vec{v}_i-\vec{v}_{i'}\|_{A_i}^{2}\right)}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{n} u_{ij}^{m} \lim_{t\to 0} \frac{\|(\vec{x}_j - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \frac{\gamma_i}{\zeta} \lim_{t\to 0} \left(\frac{1}{t} \left(\frac{1}{\frac{1}{2}\left(\|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 + \|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_{i'}}^2 \right)} \right. \right. \\
&\qquad \left. \left. - \frac{1}{\frac{1}{2}\left(\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 \right)} \right) \right) \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \frac{\gamma_{i'}}{\zeta} \lim_{t\to 0} \left(\frac{1}{t} \left(\frac{1}{\frac{1}{2}\left(\|(\vec{v}_i - \vec{v}_{i'}) + t\vec{\xi}\|_{A_{i'}}^2 + \|(\vec{v}_i - \vec{v}_{i'}) + t\vec{\xi}\|_{A_i}^2 \right)} \right. \right. \\
&\qquad \left. \left. - \frac{1}{\frac{1}{2}\left(\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 \right)} \right) \right) \\
&= \sum_{j=1}^{n} u_{ij}^{m} \lim_{t\to 0} \frac{\|\vec{x}_j - \vec{v}_i\|_{A_i}^2 - 2t(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi} - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \frac{\gamma_i}{\zeta} \lim_{t\to 0} \left(\frac{1}{t} \left(\frac{\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2}{\frac{1}{2}\left(\|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 + \|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_{i'}}^2 \right) \left(\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 \right)} \right. \right. \\
&\qquad - \left(\frac{\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 - 2t(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi}}{\frac{1}{2}\left(\|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 + \|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_{i'}}^2 \right) \left(\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 \right)} \right. \\
&\qquad \left. \left. \left. + \frac{\|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 - 2t(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_{i'} \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_{i'} \vec{\xi}}{\frac{1}{2}\left(\|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 + \|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_{i'}}^2 \right) \left(\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 \right)} \right) \right) \right) \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \frac{\gamma_{i'}}{\zeta} \lim_{t\to 0} \left(\frac{1}{t} \left(\frac{\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2}{\frac{1}{2}\left(\|(\vec{v}_i - \vec{v}_{i'}) + t\vec{\xi}\|_{A_{i'}}^2 + \|(\vec{v}_i - \vec{v}_{i'}) + t\vec{\xi}\|_{A_i}^2 \right) \left(\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 \right)} \right. \right. \\
&\qquad - \left(\frac{\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 - 2t(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_{i'} \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_{i'} \vec{\xi}}{\frac{1}{2}\left(\|(\vec{v}_i - \vec{v}_{i'}) - t\vec{\xi}\|_{A_{i'}}^2 + \|(\vec{v}_i - \vec{v}_{i'}) - t\vec{\xi}\|_{A_i}^2 \right) \left(\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 \right)} \right. \\
&\qquad \left. \left. \left. + \frac{\|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 - 2t(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi}}{\frac{1}{2}\left(\|(\vec{v}_i - \vec{v}_{i'}) - t\vec{\xi}\|_{A_{i'}}^2 + \|(\vec{v}_i - \vec{v}_{i'}) - t\vec{\xi}\|_{A_i}^2 \right) \left(\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 \right)} \right) \right) \right) \\
&= \sum_{j=1}^{n} u_j^{m} \left(-2(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} \right) + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \frac{\gamma_i}{\zeta} \left(\frac{2(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + 2(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_{i'} \vec{\xi}}{\frac{1}{2}\left(\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 \right)^2} \right) \\
&\quad - \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \frac{\gamma_{i'}}{\zeta} \left(\frac{2(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_{i'} \vec{\xi} + 2(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_i \vec{\xi}}{\frac{1}{2}\left(\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 \right)^2} \right) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
\Leftrightarrow \quad & \sum_{j=1}^{n} u_{ij}^{m} A_i^{\mathrm{T}}(\vec{x}_j - \vec{v}_i) - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_i}{\zeta} \frac{A_i^{\mathrm{T}}(\vec{v}_{i'} - \vec{v}_i) + A_{i'}^{\mathrm{T}}(\vec{v}_{i'} - \vec{v}_i)}{2d^4\left(C_i, C_{i'}\right)} \\
& + \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_{i'}}{\zeta} \frac{A_{i'}^{\mathrm{T}}(\vec{v}_i - \vec{v}_{i'}) + A_i^{\mathrm{T}}(\vec{v}_i - \vec{v}_{i'})}{2d^4\left(C_{i'}, C_i\right)} \\
= \quad & \sum_{j=1}^{n} u_{ij}^{m} A_i^{\mathrm{T}}(\vec{x}_j - \vec{v}_i) - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_i}{\zeta} \frac{A_i^{\mathrm{T}}(\vec{v}_{i'} - \vec{v}_i) + A_{i'}^{\mathrm{T}}(\vec{v}_{i'} - \vec{v}_i)}{2d^4\left(C_i, C_{i'}\right)} \\
& - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_{i'}}{\zeta} \frac{A_{i'}^{\mathrm{T}}(\vec{v}_{i'} - \vec{v}_i) + A_i^{\mathrm{T}}(\vec{v}_{i'} - \vec{v}_i)}{2d^4\left(C_i, C_{i'}\right)} \\
= \quad & 0 \\
\Leftrightarrow \quad & \sum_{j=1}^{n} u_{ij}^{m} A_i^{\mathrm{T}}(\vec{x}_j - \vec{v}_i) - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_i + \gamma_{i'}}{2\zeta d^4\left(C_i, C_{i'}\right)} \left(A_i^{\mathrm{T}}(\vec{v}_{i'} - \vec{v}_i) + A_{i'}^{\mathrm{T}}(\vec{v}_{i'} - \vec{v}_i)\right) \\
= \quad & 0 \\
\Leftrightarrow \vec{v}_i \quad = \quad & \left(\sum_{j=1}^{n} u_{ij}^{m} A_i^{\mathrm{T}} - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_i + \gamma_{i'}}{2\zeta d^4\left(C_i, C_{i'}\right)} \left(A_i^{\mathrm{T}} + A_{i'}^{\mathrm{T}}\right) \right)^{-1} \\
& \cdot \left(\sum_{j=1}^{n} u_{ij}^{m} A_i^{\mathrm{T}} \vec{x}_j - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_i + \gamma_{i'}}{2\zeta d^4\left(C_i, C_{i'}\right)} \left(A_i^{\mathrm{T}} + A_{i'}^{\mathrm{T}}\right) \vec{v}_{i'} \right)
\end{aligned}
$$

Spezialfall $A_i = A_{i'} = E$:

$$
\vec{v}_i = \frac{\displaystyle\sum_{j=1}^{n} u_{ij}^{m} \vec{x}_j - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_i + \gamma_{i'}}{\zeta d^4\left(C_i, C_{i'}\right)} \vec{v}_{i'}}{\displaystyle\sum_{j=1}^{n} u_{ij}^{m} - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_i + \gamma_{i'}}{\zeta d^4\left(C_i, C_{i'}\right)}}
$$

$$
\begin{aligned}
J_{Het1}\left(X, U, C\right) \quad = \quad & \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i) + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - u_{ij})^{m} \\
& + \sum_{i=1}^{c} \gamma_i \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{1}{\zeta \cdot \frac{1}{2} \left((\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_i (\vec{v}_{i'} - \vec{v}_i) + (\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_{i'} (\vec{v}_{i'} - \vec{v}_i) \right)} \\
& - \sum_{i=1}^{c} \lambda_i \left(\det(A_i) - 1 \right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J_{Het1}(X,U,C)}{\partial A_i} =& \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} \\
& - \frac{\gamma_i}{\zeta} \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{(\vec{v}_{i'} - \vec{v}_i)(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}}}{\frac{1}{2}\left((\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_i (\vec{v}_{i'} - \vec{v}_i) + (\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_{i'} (\vec{v}_{i'} - \vec{v}_i)\right)^2} \\
& - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\gamma_{i'}}{\zeta} \frac{(\vec{v}_i - \vec{v}_{i'})(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}}}{\frac{1}{2}\left((\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_{i'} (\vec{v}_i - \vec{v}_{i'}) + (\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_i (\vec{v}_i - \vec{v}_{i'})\right)^2} \\
& - \lambda_i \det(A_i) A_i^{-1} \\
=& 0 \\
\Leftrightarrow \lambda_i \underbrace{\det(A_i)}_{=1} A_i^{-1} =& \underbrace{\sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} - \sum_{i'=1}^{c} \frac{\gamma_i + \gamma_{i'}}{\zeta} \frac{(\vec{v}_{i'} - \vec{v}_i)(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}}}{2d^4(C_i, C_{i'})}}_{S_i} \\
\Leftrightarrow S_i =& \lambda_i A_i^{-1} \\
\Rightarrow \det(S_i A_i) =& \lambda_i^p \qquad (\text{wegen } S_i A_i = \lambda_i E) \\
\Rightarrow \lambda_i =& \sqrt[p]{\det(S_i)\det(A_i)} \\
=& \sqrt[p]{\det(S_i)} \\
\Rightarrow A_i =& \sqrt[p]{\det(S_i)} S_i^{-1}
\end{aligned}
$$

$\Rightarrow \vec{v}_i, S_i, A_i$ iterativ zu bestimmen

A.2. Herleitungen für Modellierung mittels Clusterabstoßung – Erweiterung 2

$$
J_{Het2}(X,U,C) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2(\vec{v}_i, \vec{x}_j) + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - u_{ij})^m + \sum_{i=1}^{c} \gamma_i \sum_{\substack{i'=1 \\ i' \neq i}}^{c} e^{-\zeta d^2(C_i, C_{i'})}
$$

$$
\begin{aligned}
\frac{\partial J_{Het2}(X,U,C)}{\partial \vec{v}_i} =& \sum_{j=1}^{n} u_{ij}^m \frac{\partial}{\partial \vec{v}_i} \|\vec{x}_j - \vec{v}_i\|_{A_i}^2 + \gamma_i \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{\partial}{\partial \vec{v}_i} e^{-\frac{1}{2}\zeta\left(\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2\right)} \\
& + \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \gamma_{i'} \frac{\partial}{\partial \vec{v}_i} e^{-\frac{1}{2}\zeta\left(\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2\right)}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{j=1}^{n} u_{ij}^{m} \lim_{t\to 0} \frac{\|\vec{x}_j - (\vec{v}_i + t\vec{\xi})\|_{A_i}^2 - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \gamma_i \left(-\frac{1}{2}\zeta \lim_{t\to 0} \left(\frac{1}{t} \left(\left(\|\vec{v}_{i'} - (\vec{v}_i + t\vec{\xi})\|_{A_i}^2 + \|\vec{v}_{i'} - (\vec{v}_i + t\vec{\xi})\|_{A_{i'}}^2 \right) \right.\right.\right. \\
&\qquad \left.\left.\left. - \left(\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 \right) \right) \right) \right) e^{-\frac{1}{2}\zeta \left(\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 \right)} \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \gamma_{i'} \left(-\frac{1}{2}\zeta \lim_{t\to 0} \left(\frac{1}{t} \left(\left(\|(\vec{v}_i + t\vec{\xi}) - \vec{v}_{i'}\|_{A_{i'}}^2 + \|(\vec{v}_i + t\vec{\xi}) - \vec{v}_{i'}\|_{A_i}^2 \right) \right.\right.\right. \\
&\qquad \left.\left.\left. - \left(\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 \right) \right) \right) \right) e^{-\frac{1}{2}\zeta \left(\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 \right)} \\
&= \sum_{j=1}^{n} u_{ij}^{m} \lim_{t\to 0} \frac{\|(\vec{x}_j - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \gamma_i \left(-\frac{1}{2}\zeta \lim_{t\to 0} \left(\frac{1}{t} \left(\|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 + \|(\vec{v}_{i'} - \vec{v}_i) - t\vec{\xi}\|_{A_{i'}}^2 \right.\right.\right. \\
&\qquad \left.\left.\left. - \|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 - \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 \right) \right) \right) e^{-\zeta d^2(C_i, C_{i'})} \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \gamma_{i'} \left(-\frac{1}{2}\zeta \lim_{t\to 0} \left(\frac{1}{t} \left(\|(\vec{v}_i - \vec{v}_{i'}) + t\vec{\xi}\|_{A_{i'}}^2 + \|(\vec{v}_i - \vec{v}_{i'}) + t\vec{\xi}\|_{A_i}^2 \right.\right.\right. \\
&\qquad \left.\left.\left. - \|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 - \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 \right) \right) \right) e^{-\zeta d^2(C_{i'}, C_i)} \\
&= \sum_{j=1}^{n} u_{ij}^{m} \lim_{t\to 0} \frac{\|\vec{x}_j - \vec{v}_i\|_{A_i}^2 - 2t(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi} - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \gamma_i \left(-\frac{1}{2}\zeta \lim_{t\to 0} \left(\frac{1}{t} \left(\|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 - 2t(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi} - \|\vec{v}_{i'} - \vec{v}_i\|_{A_i}^2 \right.\right.\right. \\
&\qquad \left.\left.\left. + \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 - 2t(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_{i'} \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_{i'} \vec{\xi} - \|\vec{v}_{i'} - \vec{v}_i\|_{A_{i'}}^2 \right) \right) \right) e^{-\zeta d^2(C_i, C_{i'})} \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \gamma_{i'} \left(-\frac{1}{2}\zeta \lim_{t\to 0} \left(\frac{1}{t} \left(\|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 + 2t(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_{i'} \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_{i'} \vec{\xi} - \|\vec{v}_i - \vec{v}_{i'}\|_{A_{i'}}^2 \right.\right.\right. \\
&\qquad \left.\left.\left. + \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 + 2t(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi} - \|\vec{v}_i - \vec{v}_{i'}\|_{A_i}^2 \right) \right) \right) e^{-\zeta d^2(C_{i'}, C_i)} \\
&= \sum_{j=1}^{n} u_{ij}^{m} \left(-2(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} \right) \\
&\quad + \sum_{\substack{i'=1 \\ i'\neq i}}^{c} \gamma_i \left(-\frac{1}{2}\zeta \left(-2(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} - 2(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_{i'} \vec{\xi} \right) \right) e^{-\zeta d^2(C_i, C_{i'})}
\end{aligned}
$$

$$
\begin{aligned}
& + \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \gamma_{i'} \left(-\frac{1}{2} \zeta \left(2(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_{i'} \vec{\xi} + 2(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} A_i \vec{\xi} \right) \right) e^{-\zeta d^2(C_{i'}, C_i)} \\
= \; & 0 \\
\Leftrightarrow \quad & \sum_{j=1}^{n} u_{ij}^m A_i^{\mathrm{T}} (\vec{x}_j - \vec{v}_i) - \frac{1}{2} \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \gamma_i \zeta e^{-\zeta d^2(C_i, C_{i'})} \left(A_i^{\mathrm{T}} (\vec{v}_{i'} - \vec{v}_i) + A_{i'}^{\mathrm{T}} (\vec{v}_{i'} - \vec{v}_i) \right) \\
& - \frac{1}{2} \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \gamma_{i'} \zeta e^{-\zeta d^2(C_{i'}, C_i)} \left(A_{i'}^{\mathrm{T}} (\vec{v}_{i'} - \vec{v}_i) + A_i^{\mathrm{T}} (\vec{v}_{i'} - \vec{v}_i) \right) \\
= \; & 0 \\
\Leftrightarrow \quad & \sum_{j=1}^{n} u_{ij}^m A_i^{\mathrm{T}} (\vec{x}_j - \vec{v}_i) - \frac{1}{2} \zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} (\gamma_i + \gamma_{i'}) e^{-\zeta d^2(C_i, C_{i'})} \left(A_i^{\mathrm{T}} (\vec{v}_{i'} - \vec{v}_i) + A_{i'}^{\mathrm{T}} (\vec{v}_{i'} - \vec{v}_i) \right) \\
= \; & 0 \\
\Leftrightarrow \vec{v}_i \; = \; & \left(\sum_{j=1}^{n} u_{ij}^m A_i^{\mathrm{T}} - \frac{1}{2} \zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} (\gamma_i + \gamma_{i'}) e^{-\zeta d^2(C_i, C_{i'})} \left(A_i^{\mathrm{T}} + A_{i'}^{\mathrm{T}} \right) \right)^{-1} \\
& \cdot \left(\sum_{j=1}^{n} u_{ij}^m A_i^{\mathrm{T}} \vec{x}_j - \frac{1}{2} \zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} (\gamma_i + \gamma_{i'}) e^{-\zeta d^2(C_i, C_{i'})} \left(A_i^{\mathrm{T}} + A_{i'}^{\mathrm{T}} \right) \vec{v}_{i'} \right)
\end{aligned}
$$

Spezialfall $A_i = A_{i'} = E$:

$$
\vec{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \vec{x}_j - \zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} (\gamma_i + \gamma_{i'}) e^{-\zeta d^2(C_i, C_{i'})} \vec{v}_{i'}}{\sum_{j=1}^{n} u_{ij}^m - \zeta \sum_{\substack{i'=1 \\ i' \neq i}}^{c} (\gamma_i + \gamma_{i'}) e^{-\zeta d^2(C_i, C_{i'})}}
$$

$$
\begin{aligned}
J_{Het2}(X, U, C) \; = \; & \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i) + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - u_{ij})^m \\
& + \sum_{i=1}^{c} \gamma_i \sum_{\substack{i'=1 \\ i' \neq i}}^{c} e^{-\frac{1}{2} \zeta \left((\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_i (\vec{v}_{i'} - \vec{v}_i) + (\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}} A_{i'} (\vec{v}_{i'} - \vec{v}_i) \right)} - \sum_{i=1}^{c} \lambda_i \left(\det(A_i) - 1 \right)
\end{aligned}
$$

$$
\frac{\partial J_{Het2}(X, U, C)}{\partial A_i} = \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} - \gamma_i \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{1}{2} \zeta e^{-\zeta d^2(C_i, C_{i'})} (\vec{v}_{i'} - \vec{v}_i)(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}}
$$

$$
\begin{aligned}
& - \sum_{\substack{i'=1 \\ i' \neq i}}^{c} \frac{1}{2} \gamma_{i'} \zeta e^{-\zeta d^2(C_i, C_{i'})} (\vec{v}_i - \vec{v}_{i'})(\vec{v}_i - \vec{v}_{i'})^{\mathrm{T}} - \lambda_i \det(A_i) A_i^{-1} \\
=& 0 \\
\Leftrightarrow \lambda_i \underbrace{\det(A_i)}_{=1} A_i^{-1} =& \underbrace{\sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} - \frac{1}{2} \zeta \sum_{i'=1}^{c} (\gamma_i + \gamma_{i'})\, e^{-\zeta d^2(C_i, C_{i'})} (\vec{v}_{i'} - \vec{v}_i)(\vec{v}_{i'} - \vec{v}_i)^{\mathrm{T}}}_{S_i} \\
\Leftrightarrow S_i =& \lambda_i A_i^{-1} \\
\Rightarrow \det(S_i A_i) =& \lambda_i^p \qquad (\text{wegen } S_i A_i = \lambda_i E) \\
\Rightarrow \lambda_i =& \sqrt[p]{\det(S_i) \det(A_i)} \\
=& \sqrt[p]{\det(S_i)} \\
\Rightarrow A_i =& \sqrt[p]{\det(S_i)} S_i^{-1}
\end{aligned}
$$

$\Rightarrow \vec{v}_i, S_i, A_i$ iterativ zu bestimmen

B

Herleitungen zu den Ansätzen der Clusterhomogenität

B.1. Herleitungen für Modellierung auf Basis der Clusterhomogenität – Basisansatz

$$
\begin{aligned}
J_{HomB}\left(X,U,C\right) &= \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_j\right) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(1-u_{ij}\right)^{m} \\
&\quad + \frac{1}{2}\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\left(d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_k\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J_{HomB}\left(X,U,C\right)}{\partial \vec{v}_i} &= \sum_{j=1}^{n} u_{ij}^{m}\frac{\partial}{\partial \vec{v}_i}\|\vec{x}_j-\vec{v}_i\|_{A_i}^{2} + \frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\left(\frac{\partial}{\partial \vec{v}_i}\|\vec{x}_j-\vec{v}_i\|_{A_i}^{2}+\frac{\partial}{\partial \vec{v}_i}\|\vec{x}_k-\vec{v}_i\|_{A_i}^{2}\right) \\
&= \sum_{j=1}^{m} u_{ij}^{m}\lim_{t\to 0}\frac{\|\vec{x}_j-(\vec{v}_i+t\vec{\xi})\|_{A_i}^{2}-\|\vec{x}_j-\vec{v}_i\|_{A_i}^{2}}{t} \\
&\quad + \frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\left(\lim_{t\to 0}\frac{\|\vec{x}_j-(\vec{v}_i+t\vec{\xi})\|_{A_i}^{2}-\|\vec{x}_j-\vec{v}_i\|_{A_i}^{2}}{t}\right. \\
&\qquad \left. + \lim_{t\to 0}\frac{\|\vec{x}_k-(\vec{v}_i+t\vec{\xi})\|_{A_i}^{2}-\|\vec{x}_k-\vec{v}_i\|_{A_i}^{2}}{t}\right) \\
&= \sum_{j=1}^{n} u_{ij}^{m}\lim_{t\to 0}\frac{\|(\vec{x}_j-\vec{v}_i)-t\vec{\xi}\|_{A_i}^{2}-\|\vec{x}_j-\vec{v}_i\|_{A_i}^{2}}{t} \\
&\quad + \frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\left(\lim_{t\to 0}\frac{\|(\vec{x}_j-\vec{v}_i)-t\vec{\xi}\|_{A_i}^{2}-\|\vec{x}_j-\vec{v}_i\|_{A_i}^{2}}{t}\right.
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \lim_{t\to 0} \frac{\|(\vec{x}_k - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 - \|\vec{x}_k - \vec{v}_i\|_{A_i}^2}{t} \Bigg) \\
&= \sum_{j=1}^{n} u_{ij}^m \lim_{t\to 0} \frac{\|\vec{x}_j - \vec{v}_i\|_{A_i}^2 - 2t(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi} - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \Bigg(\lim_{t\to 0} \frac{\|\vec{x}_j - \vec{v}_i\|_{A_i}^2 - 2t(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi} - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \lim_{t\to 0} \frac{\|\vec{x}_k - \vec{v}_i\|_{A_i}^2 - 2t(\vec{x}_k - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi} - \|\vec{x}_k - \vec{v}_i\|_{A_i}^2}{t} \Bigg) \\
&= \sum_{j=1}^{n} u_{ij}^m \left(-2(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} \right) \\
&\quad + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \left(-2(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} - 2(\vec{x}_k - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} \right) \\
&= 0 \\
\Leftrightarrow \quad & \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i) + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} ((\vec{x}_j - \vec{v}_i) + (\vec{x}_k - \vec{v}_i)) \\
&= 0 \\
& \sum_{j=1}^{n} u_{ij}^m \vec{x}_j + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} (\vec{x}_j + \vec{x}_k) \\
\Leftrightarrow \vec{v}_i &= \frac{}{\sum_{j=1}^{n} u_{ij}^m + \frac{\gamma_i}{2\zeta_i} n_i^\alpha \left(n_i^\alpha - 1\right)}
\end{aligned}
$$

$$
\begin{aligned}
J_{HomB}(X,U,C) &= \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i) + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - u_{ij})^m \\
&\quad + \frac{1}{2} \sum_{i=1}^{c} \frac{\gamma_i}{\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} ((\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i) + (\vec{x}_k - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_k - \vec{v}_i)) \\
&\quad - \sum_{i=1}^{c} \lambda_i \left(\det(A_i) - 1\right)
\end{aligned}
$$

$$
\frac{\partial J_{HomB}(X,U,C)}{\partial A_i} = \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}}
$$

$$
\begin{aligned}
&+\frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\k>j}}((\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}+(\vec{x}_k-\vec{v}_i)(\vec{x}_k-\vec{v}_i)^{\mathrm{T}})\\
&-\lambda_i\det(A_i)A_i^{-1}\\
=&0\\
\Leftrightarrow\lambda_i\underbrace{\det(A_i)}_{=1}A_i^{-1}=&\sum_{j=1}^{n}u_{ij}^m(\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}\\
&+\frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\k>j}}\left((\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}+(\vec{x}_k-\vec{v}_i)(\vec{x}_k-\vec{v}_i)^{\mathrm{T}}\right)\\
=&S_i\\
\Rightarrow\det(S_iA_i)=&\lambda_i^p\qquad(\text{wegen } S_iA_i=\lambda_iE)\\
\Rightarrow\lambda_i=&\sqrt[p]{\det(S_i)\det(A_i)}\\
=&\sqrt[p]{\det(S_i)}\\
\Rightarrow A_i=&\sqrt[p]{\det(S_i)}S_i^{-1}
\end{aligned}
$$

$\Rightarrow S_i, A_i$ iterativ zu bestimmen

B.2. Herleitungen für Modellierung auf Basis der Clusterhomogenität – Verhältnis der Distanzen

$$
\begin{aligned}
J_{HomV1}(X,U,C) &= \sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m d_{C_i}^2(\vec{v}_i,\vec{x}_j)+\sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}(1-u_{ij})^m\\
&+\frac{1}{2}\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\k>j}}\frac{d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)}{d_{C_i}^2(\vec{x}_j,\vec{x}_k)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J_{HomV1}(X,U,C)}{\partial\vec{v}_i} &= \sum_{j=1}^{n}u_{ij}^m\frac{\partial}{\partial\vec{v}_i}\|\vec{x}_j-\vec{v}_i\|_{A_i}^2\\
&+\frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\k>j}}\frac{1}{d_{C_i}^2(\vec{x}_j,\vec{x}_k)}\left(\frac{\partial}{\partial\vec{v}_i}\|\vec{x}_j-\vec{v}_i\|_{A_i}^2+\frac{\partial}{\partial\vec{v}_i}\|\vec{x}_k-\vec{v}_i\|_{A_i}^2\right)\\
&= \sum_{j=1}^{m}u_{ij}^m\lim_{t\to0}\frac{\|\vec{x}_j-(\vec{v}_i+t\vec{\xi})\|_{A_i}^2-\|\vec{x}_j-\vec{v}_i\|_{A_i}^2}{t}\\
&+\frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\k>j}}\frac{1}{d_{C_i}^2(\vec{x}_j,\vec{x}_k)}\left(\lim_{t\to0}\frac{\|\vec{x}_j-(\vec{v}_i+t\vec{\xi})\|_{A_i}^2-\|\vec{x}_j-\vec{v}_i\|_{A_i}^2}{t}\right.
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \lim_{t\to 0} \frac{\|\vec{x}_k - (\vec{v}_i + t\vec{\xi})\|_{A_i}^2 - \|\vec{x}_k - \vec{v}_i\|_{A_i}^2}{t} \Bigg) \\
&= \sum_{j=1}^{n} u_{ij}^m \lim_{t\to 0} \frac{\|(\vec{x}_j - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \frac{1}{d_{C_i}^2(\vec{x}_j, \vec{x}_k)} \Bigg(\lim_{t\to 0} \frac{\|(\vec{x}_j - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \lim_{t\to 0} \frac{\|(\vec{x}_k - \vec{v}_i) - t\vec{\xi}\|_{A_i}^2 - \|\vec{x}_k - \vec{v}_i\|_{A_i}^2}{t} \Bigg) \\
&= \sum_{j=1}^{n} u_{ij}^m \lim_{t\to 0} \frac{\|\vec{x}_j - \vec{v}_i\|_{A_i}^2 - 2t(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi} - \|\vec{x}_j - \vec{v}_i\|_{A_i}^2}{t} \\
&\quad + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \frac{1}{d_{C_i}^2(\vec{x}_j, \vec{x}_k)} \Bigg(\lim_{t\to 0} \Bigg(\frac{\|\vec{x}_j - \vec{v}_i\|_{A_i}^2 - 2t(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi}}{t} \\
&\qquad - \frac{\|\vec{x}_j \vec{v}_i\|_{A_i}^2}{t} \Bigg) \\
&\quad + \lim_{t\to 0} \frac{\|\vec{x}_k - \vec{v}_i\|_{A_i}^2 - 2t(\vec{x}_k - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} + t^2 \vec{\xi}^{\mathrm{T}} A_i \vec{\xi} - \|\vec{x}_k - \vec{v}_i\|_{A_i}^2}{t} \Bigg) \\
&= \sum_{j=1}^{n} u_{ij}^m \left(-2(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} \right) \\
&\quad + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \frac{1}{d_{C_i}^2(\vec{x}_j, \vec{x}_k)} \left(-2(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} - 2(\vec{x}_k - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} \right) \\
&= 0 \\
\Leftrightarrow &\quad \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i) + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \frac{1}{d_{C_i}^2(\vec{x}_j, \vec{x}_k)} \left((\vec{x}_j - \vec{v}_i) + (\vec{x}_k - \vec{v}_i) \right) \\
&= 0 \\
&\quad \sum_{j=1}^{n} u_{ij}^m \vec{x}_j + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \frac{1}{d_{C_i}^2(\vec{x}_j, \vec{x}_k)} (\vec{x}_j + \vec{x}_k) \\
\Leftrightarrow \vec{v}_i &= \frac{}{\sum_{j=1}^{n} u_{ij}^m + \frac{\gamma_i}{2\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \frac{1}{d_{C_i}^2(\vec{x}_j, \vec{x}_k)}}
\end{aligned}
$$

$$
\begin{aligned}
J_{HomV1}(X,U,C) &= \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}A_i(\vec{x}_j-\vec{v}_i)+\sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}(1-u_{ij})^m \\
&\quad +\frac{1}{2}\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\ k>j}}\frac{(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}A_i(\vec{x}_j-\vec{v}_i)+(\vec{x}_k-\vec{v}_i)^{\mathrm{T}}A_i(\vec{x}_k-\vec{v}_i)}{(\vec{x}_k-\vec{x}_j)^{\mathrm{T}}(\vec{x}_k-\vec{x}_j)} \\
&\quad -\sum_{i=1}^{c}\lambda_i\left(\det(A_i)-1\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J_{HomV1}(X,U,C)}{\partial A_i} &= \sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} \\
&\quad +\frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\ k>j}}\left(\frac{d_{C_i}^2(\vec{x}_j,\vec{x}_k)\left((\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}+(\vec{x}_k-\vec{v}_i)(\vec{x}_k-\vec{v}_i)^{\mathrm{T}}\right)}{d_{C_i}^4(\vec{x}_j,\vec{x}_k)}\right. \\
&\qquad \left.-\frac{\left(d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)\right)(\vec{x}_k-\vec{x}_j)(\vec{x}_k-\vec{x}_j)^{\mathrm{T}}}{d_{C_i}^4(\vec{x}_j,\vec{x}_k)}\right) \\
&\quad -\lambda_i\det(A_i)A_i^{-1} \\
&= 0 \\
\Leftrightarrow \lambda_i\underbrace{\det(A_i)}_{=1}A_i^{-1} &= \sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} \\
&\quad +\frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\ k>j}}\left(\frac{d_{C_i}^2(\vec{x}_j,\vec{x}_k)\left((\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}+(\vec{x}_k-\vec{v}_i)(\vec{x}_k-\vec{v}_i)^{\mathrm{T}}\right)}{d_{C_i}^4(\vec{x}_j,\vec{x}_k)}\right. \\
&\qquad \left.-\frac{\left(d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)\right)(\vec{x}_k-\vec{x}_j)(\vec{x}_k-\vec{x}_j)^{\mathrm{T}}}{d_{C_i}^4(\vec{x}_j,\vec{x}_k)}\right) \\
&= S_i \\
\Leftrightarrow S_i &= \lambda_i A_i^{-1} \\
\Rightarrow \det(S_iA_i) &= \lambda_i^p \qquad (\text{wegen } S_iA_i=\lambda_iE) \\
\Rightarrow \lambda_i &= \sqrt[p]{\det(S_i)\det(A_i)} \\
&= \sqrt[p]{\det(S_i)} \\
\Rightarrow A_i &= \sqrt[p]{\det(S_i)}S_i^{-1}
\end{aligned}
$$

$\Rightarrow S_i, A_i$ iterativ zu bestimmen

$$
\begin{aligned}
J_{HomV2}\left(X,U,C\right) &= \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(1-u_{ij}\right)^m \\
&\quad +2\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\frac{d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)}{d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J_{HomV2}\left(X,U,C\right)}{\partial \vec{v}_i} &= \sum_{j=1}^{n} u_{ij}^m \frac{\partial}{\partial \vec{v}_i}\|\vec{x}_j-\vec{v}_i\|_{A_i}^2 \\
&\quad +\frac{2\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\frac{\partial}{\partial\vec{v}_i}\frac{d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)}{\|\vec{x}_j-\vec{v}_i\|_{A_i}^2+\|\vec{x}_k-\vec{v}_i\|_{A_i}^2} \\
&= \text{... (vgl. Herleitung für HomV1)} \\
&= \sum_{j=1}^{n} u_{ij}^m\left(-2(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}A_i\vec{\xi}\right) \\
&\quad -\frac{2\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\frac{d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)\left(-2\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}A_i\vec{\xi}-2\left(\vec{x}_k-\vec{v}_i\right)^{\mathrm{T}}A_i\vec{\xi}\right)}{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)^2} \\
&= 0 \\
\Leftrightarrow \quad & \sum_{j=1}^{n} u_{ij}^m(\vec{x}_j-\vec{v}_i) - \frac{2\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\frac{d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)\left(\left(\vec{x}_j-\vec{v}_i\right)+\left(\vec{x}_k-\vec{v}_i\right)\right)}{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)^2} \\
&= 0 \\
& \sum_{j=1}^{n} u_{ij}^m\vec{x}_j - \frac{2\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\frac{d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)}{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)^2}\left(\vec{x}_j+\vec{x}_k\right) \\
\Leftrightarrow \vec{v}_i &= \frac{}{\sum_{j=1}^{n} u_{ij}^m - \frac{4\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\frac{d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)}{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)^2}}
\end{aligned}
$$

$$
\begin{aligned}
J_{HomV2}\left(X,U,C\right) &= \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}A_i(\vec{x}_j-\vec{v}_i) + \sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}\left(1-u_{ij}\right)^m \\
&\quad +2\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\ k>j}}\frac{\left(\vec{x}_k-\vec{x}_j\right)^{\mathrm{T}}A_i\left(\vec{x}_k-\vec{x}_j\right)}{\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}A_i\left(\vec{x}_j-\vec{v}_i\right)+\left(\vec{x}_k-\vec{v}_i\right)^{\mathrm{T}}A_i\left(\vec{x}_k-\vec{v}_i\right)} \\
&\quad -\sum_{i=1}^{c}\lambda_i\left(\det(A_i)-1\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J_{HomV2}\left(X,U,C\right)}{\partial A_i} &= \sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} \\
&\quad +\frac{2\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\ k>j}}\left(\frac{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)\left(\vec{x}_k-\vec{x}_j\right)\left(\vec{x}_k-\vec{x}_j\right)^{\mathrm{T}}}{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)^2}\right. \\
&\qquad \left.-\frac{d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)\left(\left(\vec{x}_j-\vec{v}_i\right)\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}+\left(\vec{x}_k-\vec{v}_i\right)\left(\vec{x}_k-\vec{v}_i\right)^{\mathrm{T}}\right)}{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)^2}\right) \\
&\quad -\lambda_i\det(A_i)A_i^{-1} \\
&= 0 \\
\Leftrightarrow \lambda_i\underbrace{\det(A_i)}_{=1}A_i^{-1} &= \sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} \\
&\quad +\frac{2\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\ k>j}}\left(\frac{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)\left(\vec{x}_k-\vec{x}_j\right)\left(\vec{x}_k-\vec{x}_j\right)^{\mathrm{T}}}{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)^2}\right. \\
&\qquad \left.-\frac{d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)\left(\left(\vec{x}_j-\vec{v}_i\right)\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}+\left(\vec{x}_k-\vec{v}_i\right)\left(\vec{x}_k-\vec{v}_i\right)^{\mathrm{T}}\right)}{\left(d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^2\left(\vec{v}_i,\vec{x}_k\right)\right)^2}\right) \\
&= S_i \\
\Leftrightarrow S_i &= \lambda_i A_i^{-1} \\
\Rightarrow \det(S_iA_i) &= \lambda_i^p \qquad (\text{wegen } S_iA_i=\lambda_iE) \\
\Rightarrow \lambda_i &= \sqrt[p]{\det(S_i)\det(A_i)} \\
&= \sqrt[p]{\det(S_i)} \\
\Rightarrow A_i &= \sqrt[p]{\det(S_i)}S_i^{-1}
\end{aligned}
$$

$\Rightarrow \vec{v}_i, S_i, A_i$ iterativ zu bestimmen

$$
\begin{aligned}
J_{HomV3}\left(X,U,C\right) &= \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(1-u_{ij}\right)^m \\
&\quad + \sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i} n_i^{\alpha} \frac{\sum\limits_{\vec{x}_j\in[U_i]_\alpha}\sum\limits_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)}{\sum\limits_{\vec{x}_j\in[U_i]_\alpha} d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J_{HomV3}\left(X,U,C\right)}{\partial \vec{v}_i} &= \sum_{j=1}^{n} u_{ij}^{m} \frac{\partial}{\partial \vec{v}_i}\|\vec{x}_j-\vec{v}_i\|_{A_i}^2 \\
&\quad + \frac{\gamma_i}{\zeta_i} n_i^{\alpha} \frac{\partial}{\partial \vec{v}_i} \frac{\sum\limits_{\vec{x}_j\in[U_i]_\alpha}\sum\limits_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)}{\sum\limits_{\vec{x}_j\in[U_i]_\alpha}\|\vec{x}_j-\vec{v}_i\|_{A_i}^2} \\
&= \text{... (vgl. Herleitung für HomV1)} \\
&= \sum_{j=1}^{n} u_{ij}^{m}\left(-2(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} A_i \vec{\xi}\right) \\
&\quad - \frac{\gamma_i}{\zeta_i} n_i^{\alpha} \frac{\left(\sum\limits_{\vec{x}_j\in[U_i]_\alpha}\sum\limits_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)\right)\left(\sum\limits_{\vec{x}_j\in[U_i]_\alpha}\left(-2\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}} A_i\vec{\xi}\right)\right)}{\left(\sum\limits_{\vec{x}_j\in[U_i]_\alpha} d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)\right)^2} \\
&= 0 \\
\Leftrightarrow \quad & \sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i) - \frac{\gamma_i}{\zeta_i} n_i^{\alpha} \frac{\left(\sum\limits_{\vec{x}_j\in[U_i]_\alpha}\sum\limits_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2\left(\vec{x}_j,\vec{x}_k\right)\right)\left(\sum\limits_{\vec{x}_j\in[U_i]_\alpha}\left(\vec{x}_j-\vec{v}_i\right)\right)}{\left(\sum\limits_{\vec{x}_j\in[U_i]_\alpha} d_{C_i}^2\left(\vec{v}_i,\vec{x}_j\right)\right)^2} \\
&= 0
\end{aligned}
$$

$$\Leftrightarrow \vec{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \vec{x}_j - \frac{\gamma_i}{\zeta_i} n_i^\alpha \frac{\left(\sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} d_{C_i}^2 (\vec{x}_j, \vec{x}_k)\right)}{\left(\sum_{\vec{x}_j \in [U_i]_\alpha} d_{C_i}^2 (\vec{v}_i, \vec{x}_j)\right)^2} \left(\sum_{\vec{x}_j \in [U_i]_\alpha} \vec{x}_j\right)}{\sum_{j=1}^{n} u_{ij}^m - \frac{\gamma_i}{\zeta_i} n_i^{\alpha 2} \frac{\left(\sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} d_{C_i}^2 (\vec{x}_j, \vec{x}_k)\right)}{\left(\sum_{\vec{x}_j \in [U_i]_\alpha} d_{C_i}^2 (\vec{v}_i, \vec{x}_j)\right)^2}}$$

$$\begin{aligned} J_{HomV3}(X, U, C) = & \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i) + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - u_{ij})^m \\ & + \sum_{i=1}^{c} \frac{\gamma_i}{\zeta_i} n_i^\alpha \frac{\sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} (\vec{x}_k - \vec{x}_j)^{\mathrm{T}} A_i (\vec{x}_k - \vec{x}_j)}{\sum_{\vec{x}_j \in [U_i]_\alpha} (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i)} \\ & - \sum_{i=1}^{c} \lambda_i (\det(A_i) - 1) \end{aligned}$$

$$\begin{aligned} \frac{\partial J_{HomV3}(X, U, C)}{\partial A_i} = & \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} \\ & + \frac{\gamma_i}{\zeta_i} n_i^\alpha \left(\frac{\left(\sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} (\vec{x}_k - \vec{x}_j)(\vec{x}_k - \vec{x}_j)^{\mathrm{T}} \right) \left(\sum_{\vec{x}_j \in [U_i]_\alpha} d_{C_i}^2 (\vec{v}_i, \vec{x}_j) \right)}{\left(\sum_{\vec{x}_j \in [U_i]_\alpha} (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i) \right)^2} \right. \\ & \left. - \frac{\left(\sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} d_{C_i}^2 (\vec{x}_j, \vec{x}_k) \right) \left(\sum_{\vec{x}_j \in [U_i]_\alpha} (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} \right)}{\left(\sum_{\vec{x}_j \in [U_i]_\alpha} (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i) \right)^2} \right) \end{aligned}$$

$$
\begin{aligned}
& - \lambda_i \det(A_i) A_i^{-1} \\
=& 0 \\
\Leftrightarrow \lambda_i \underbrace{\det(A_i)}_{=1} A_i^{-1} =& \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} \\
& + \frac{\gamma_i}{\zeta_i} n_i^{\alpha} \left(\frac{\left(\sum\limits_{\vec{x}_j \in [U_i]_\alpha} \sum\limits_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} (\vec{x}_k - \vec{x}_j)(\vec{x}_k - \vec{x}_j)^{\mathrm{T}} \right) \left(\sum\limits_{\vec{x}_j \in [U_i]_\alpha} d_{C_i}^2(\vec{v}_i, \vec{x}_j) \right)}{\left(\sum\limits_{\vec{x}_j \in [U_i]_\alpha} (\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i (\vec{x}_j - \vec{v}_i) \right)^2} \right. \\
& \left. - \frac{\left(\sum\limits_{\vec{x}_j \in [U_i]_\alpha} \sum\limits_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} d_{C_i}^2(\vec{x}_j, \vec{x}_k) \right) \left(\sum\limits_{\vec{x}_j \in [U_i]_\alpha} (\vec{x}_j - \vec{v}_i)(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} \right)}{\left(\sum\limits_{\vec{x}_j \in [U_i]_\alpha} d_{C_i}^2(\vec{v}_i, \vec{x}_j) \right)^2} \right) \\
=& S_i \\
\Leftrightarrow S_i =& \lambda_i A_i^{-1} \\
\Rightarrow \det(S_i A_i) =& \lambda_i^p \qquad (\text{wegen } S_i A_i = \lambda_i E) \\
\Rightarrow \lambda_i =& \sqrt[p]{\det(S_i)\det(A_i)} \\
=& \sqrt[p]{\det(S_i)} \\
\Rightarrow A_i =& \sqrt[p]{\det(S_i)} S_i^{-1}
\end{aligned}
$$

$\Rightarrow \vec{v}_i, S_i, A_i$ iterativ zu bestimmen

$$
\begin{aligned}
J_{HomV4}(X,U,C) =& \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{C_i}^2(\vec{v}_i, \vec{x}_j) + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - u_{ij})^m \\
& + \sum_{i=1}^{c} \frac{\gamma_i}{\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \left(1 - \frac{2 d_{C_i}^2(\vec{x}_j, \vec{x}_k)}{d_{C_i}^2(\vec{v}_i, \vec{x}_j) + d_{C_i}^2(\vec{v}_i, \vec{x}_k)} \right)^2
\end{aligned}
$$

$$
\frac{\partial J_{HomV4}(X,U,C)}{\partial \vec{v}_i} = \sum_{j=1}^{n} u_{ij}^m \frac{\partial}{\partial \vec{v}_i} \|\vec{x}_j - \vec{v}_i\|_{A_i}^2
$$

$$
\begin{aligned}
&+\frac{\gamma_i}{\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \frac{\partial}{\partial \vec{v}_i} \left(1 - \frac{2d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{\|\vec{x}_j - \vec{v}_i\|^2_{A_i} + \|\vec{x}_k - \vec{v}_i\|^2_{A_i}}\right)^2 \\
= \;& \text{... (vgl. Herleitung für HomV1)} \\
= \;& \sum_{j=1}^{n} u_{ij}^m \left(-2(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi}\right) \\
&+\frac{4\gamma_i}{\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \left(\left(1 - \frac{2d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)}\right)\right. \\
&\qquad \left. \cdot \frac{d^2_{C_i}(\vec{x}_j, \vec{x}_k)\left(-2(\vec{x}_j - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi} - 2(\vec{x}_k - \vec{v}_i)^{\mathrm{T}} A_i \vec{\xi}\right)}{\left(d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)\right)^2}\right) \\
= \;& 0 \\
\Leftrightarrow \quad & \sum_{j=1}^{n} u_{ij}^m (\vec{x}_j - \vec{v}_i) \\
&+\frac{4\gamma_i}{\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \left(\left(1 - \frac{2d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)}\right)\right. \\
&\qquad \left. \cdot \frac{d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{\left(d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)\right)^2} ((\vec{x}_j - \vec{v}_i) + (\vec{x}_k - \vec{v}_i))\right) \\
= \;& 0 \\
\Leftrightarrow \vec{v}_i \;=\; & \left(\sum_{j=1}^{n} u_{ij}^m + \frac{8\gamma_i}{\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \left(\left(1 - \frac{d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)}\right)\right.\right. \\
&\qquad \left.\left. \cdot \frac{d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{\left(d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)\right)^2}\right)\right)^{-1} \\
&\cdot \left(\sum_{j=1}^{n} u_{ij}^m \vec{x}_j + \frac{4\gamma_i}{\zeta_i} \sum_{\vec{x}_j \in [U_i]_\alpha} \sum_{\substack{\vec{x}_k \in [U_i]_\alpha \\ k>j}} \left(\left(1 - \frac{d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)}\right)\right.\right. \\
&\qquad \left.\left. \cdot \frac{d^2_{C_i}(\vec{x}_j, \vec{x}_k)}{\left(d^2_{C_i}(\vec{v}_i, \vec{x}_j) + d^2_{C_i}(\vec{v}_i, \vec{x}_k)\right)^2} (\vec{x}_j + \vec{x}_k)\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
J_{HomV4}(X,U,C) &= \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}A_i(\vec{x}_j-\vec{v}_i)+\sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}(1-u_{ij})^m \\
&\quad+\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\ k>j}}\left(1-\frac{2\left(\vec{x}_k-\vec{x}_j\right)^{\mathrm{T}}A_i\left(\vec{x}_k-\vec{x}_j\right)}{\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}A_i\left(\vec{x}_j-\vec{v}_i\right)+\left(\vec{x}_k-\vec{v}_i\right)^{\mathrm{T}}A_i\left(\vec{x}_k-\vec{v}_i\right)}\right)^2 \\
&\quad-\sum_{i=1}^{c}\lambda_i\left(\det(A_i)-1\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J_{HomV4}(X,U,C)}{\partial A_i} &= \sum_{j=1}^{n}u_{ij}^{m}(\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} \\
&\quad-\frac{4\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\ k>j}}\left(\left(1-\frac{2d_{C_i}^2(\vec{x}_j,\vec{x}_k)}{d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)}\right)\right. \\
&\qquad\cdot\left(\frac{\left(\vec{x}_k-\vec{x}_j\right)\left(\vec{x}_k-\vec{x}_j\right)^{\mathrm{T}}\left(d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)\right)}{\left(d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)\right)^2}\right. \\
&\qquad\left.\left.-\frac{d_{C_i}^2(\vec{x}_j,\vec{x}_k)\left(\left(\vec{x}_j-\vec{v}_i\right)\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}+\left(\vec{x}_k-\vec{v}_i\right)\left(\vec{x}_k-\vec{v}_i\right)^{\mathrm{T}}\right)}{\left(d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)\right)^2}\right)\right) \\
&\quad-\lambda_i\det(A_i)A_i^{-1} \\
&=0 \\
\Leftrightarrow\lambda_i\underbrace{\det(A_i)}_{=1}A_i^{-1} &= \sum_{j=1}^{n}u_{ij}^{m}(\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} \\
&\quad-\frac{4\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha\\ k>j}}\left(\left(1-\frac{2d_{C_i}^2(\vec{x}_j,\vec{x}_k)}{d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)}\right)\right. \\
&\qquad\cdot\left(\frac{\left(\vec{x}_k-\vec{x}_j\right)\left(\vec{x}_k-\vec{x}_j\right)^{\mathrm{T}}\left(d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)\right)}{\left(d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)\right)^2}\right. \\
&\qquad\left.\left.-\frac{d_{C_i}^2(\vec{x}_j,\vec{x}_k)\left(\left(\vec{x}_j-\vec{v}_i\right)\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}+\left(\vec{x}_k-\vec{v}_i\right)\left(\vec{x}_k-\vec{v}_i\right)^{\mathrm{T}}\right)}{\left(d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)\right)^2}\right)\right) \\
&=S_i \\
\Leftrightarrow S_i &=\lambda_i A_i^{-1} \\
\Rightarrow\det(S_iA_i) &=\lambda_i^p \qquad (\text{wegen } S_iA_i=\lambda_iE) \\
\Rightarrow\lambda_i &=\sqrt[p]{\det(S_i)\det(A_i)} \\
&=\sqrt[p]{\det(S_i)}
\end{aligned}
$$

$$\Rightarrow A_i = \sqrt[p]{\det(S_i)} S_i^{-1}$$

$\Rightarrow \vec{v}_i, S_i, A_i$ iterativ zu bestimmen

B.3. Herleitungen für Modellierung auf Basis der Clusterhomogenität – Dreiecksbeziehung der Distanzen

$$\begin{aligned} J_{HomD}(X,U,C) &= \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{C_i}^2(\vec{v}_i,\vec{x}_j) + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}(1-u_{ij})^m \\ &\quad + \frac{1}{2}\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2(\vec{x}_j,\vec{x}_k)\left(d_{C_i}^2(\vec{v}_i,\vec{x}_j)+d_{C_i}^2(\vec{v}_i,\vec{x}_k)\right) \end{aligned}$$

$$\begin{aligned} \frac{\partial J_{HomD}(X,U,C)}{\partial \vec{v}_i} &= \sum_{j=1}^{n} u_{ij}^m \frac{\partial}{\partial \vec{v}_i}\|\vec{x}_j-\vec{v}_i\|_{A_i}^2 \\ &\quad + \frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2\left(\frac{\partial}{\partial \vec{v}_i}\|\vec{x}_j-\vec{v}_i\|_{A_i}^2 + \frac{\partial}{\partial \vec{v}_i}\|\vec{x}_k-\vec{v}_i\|_{A_i}^2\right) \\ &= \text{... (vgl. Herleitung für HomB)} \\ &= \sum_{j=1}^{n} u_{ij}^m\left(-2(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} A_i\vec{\xi}\right) \\ &\quad + \frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2(\vec{x}_j,\vec{x}_k)\left(-2(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} A_i\vec{\xi} - 2(\vec{x}_k-\vec{v}_i)^{\mathrm{T}} A_i\vec{\xi}\right) \\ &= 0 \\ \Leftrightarrow &\quad \sum_{j=1}^{n} u_{ij}^m(\vec{x}_j-\vec{v}_i) + \frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2(\vec{x}_j,\vec{x}_k)\left((\vec{x}_j-\vec{v}_i)+(\vec{x}_k-\vec{v}_i)\right) \\ &= 0 \\ &\quad \sum_{j=1}^{n} u_{ij}^m\vec{x}_j + \frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2(\vec{x}_j,\vec{x}_k)(\vec{x}_j+\vec{x}_k) \\ \Leftrightarrow \vec{v}_i &= \frac{}{\sum_{j=1}^{n} u_{ij}^m + \frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}} d_{C_i}^2(\vec{x}_j,\vec{x}_k)} \end{aligned}$$

$$
\begin{aligned}
J_{HomD}\left(X,U,C\right) \;=\; & \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)^{\mathrm{T}}A_i(\vec{x}_j-\vec{v}_i)+\sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}\left(1-u_{ij}\right)^{m} \\
& +\frac{1}{2}\sum_{i=1}^{c}\frac{\gamma_i}{\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\left(\vec{x}_k-\vec{x}_j\right)^{\mathrm{T}}A_i\left(\vec{x}_j-\vec{x}_k\right)\Big(\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}A_i\left(\vec{x}_j-\vec{v}_i\right) \\
& \qquad +\left(\vec{x}_k-\vec{v}_i\right)^{\mathrm{T}}A_i\left(\vec{x}_k-\vec{v}_i\right)\Big) \\
& -\sum_{i=1}^{c}\lambda_i\left(\det(A_i)-1\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J_{HomD}\left(X,U,C\right)}{\partial A_i} =& \sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} \\
& +\frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\Big(\left(\vec{x}_k-\vec{x}_j\right)\left(\vec{x}_k-\vec{x}_j\right)^{\mathrm{T}}\left(d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_k\right)\right) \\
& \qquad +d_{C_i}^{2}\left(\vec{x}_k,\vec{x}_k\right)\Big(\left(\vec{x}_j-\vec{v}_i\right)\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}+\left(\vec{x}_k-\vec{v}_i\right)\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}\Big)\Big) \\
& -\lambda_i\det(A_i)A_i^{-1} \\
=& 0 \\
\Leftrightarrow \lambda_i\underbrace{\det(A_i)}_{=1}A_i^{-1} =& \sum_{j=1}^{n} u_{ij}^{m}(\vec{x}_j-\vec{v}_i)(\vec{x}_j-\vec{v}_i)^{\mathrm{T}} \\
& +\frac{\gamma_i}{2\zeta_i}\sum_{\vec{x}_j\in[U_i]_\alpha}\sum_{\substack{\vec{x}_k\in[U_i]_\alpha \\ k>j}}\Big(\left(\vec{x}_k-\vec{x}_j\right)\left(\vec{x}_k-\vec{x}_j\right)^{\mathrm{T}}\left(d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_j\right)+d_{C_i}^{2}\left(\vec{v}_i,\vec{x}_k\right)\right) \\
& \qquad +d_{C_i}^{2}\left(\vec{x}_j,\vec{x}_k\right)\Big(\left(\vec{x}_j-\vec{v}_i\right)\left(\vec{x}_j-\vec{v}_i\right)^{\mathrm{T}}+\left(\vec{x}_k-\vec{v}_i\right)\left(\vec{x}_k-\vec{v}_i\right)^{\mathrm{T}}\Big)\Big) \\
=& S_i \\
\Leftrightarrow S_i =& \lambda_i A_i^{-1} \\
\Rightarrow \det(S_iA_i) =& \lambda_i^{p} \qquad (\text{wegen } S_iA_i=\lambda_iE) \\
\Rightarrow \lambda_i =& \sqrt[p]{\det(S_i)\det(A_i)} \\
=& \sqrt[p]{\det(S_i)} \\
\Rightarrow A_i =& \sqrt[p]{\det(S_i)}S_i^{-1}
\end{aligned}
$$

$\Rightarrow S_i, A_i$ iterativ zu bestimmen

C

Erläuterungen zu den durchgeführten Experimenten

Für die in Kapitel 5 durchgeführten Experimente wurde ein Java-Applet entwickelt, das eine orts- und plattformunabhängige Durchführung der Experimente ermöglicht und die für die Analyse benötigten Funktionen enthält. In den folgenden Abschnitten werden der Aufbau und die Funktionalitäten des Applets, die Datenausgabe sowie die Durchführung der einzelnen Experimente detailliert erläutert.

C.1. Aufbau und Funktionalitäten des Java-Applets

Das Applet übernimmt zwei wesentliche Funktionsbereiche für die Durchführung verschiedener Experimente, die Datengenerierung und die Fuzzy-Clusteranalyse, nach denen der Aufbau des Applets gegliedert ist. Bevor die eigentliche Analyse im dynamischen Kontext durchgeführt wird, erfolgt zunächst eine Eingabe der benötigten Parameter. Die Ergebnisse der Analyse werden anschließend in einem einfachen Ausgabefenster dargestellt.

C.1.1. Dateneingabe

Neben den Parametern für die Generierung der Daten werden verschiedene Analyseparameter benötigt, die im Folgenden stichpunktartig aufgeführt werden.

Parameter zur Datengenerierung

Zu Beginn werden die allgemeinen Parameter bestimmt:

- *Anzahl Cluster für Datengenerierung*: Die Clusterzahl bei der Datengenerierung muss nicht zwangsläufig der für die Analyse festgesetzten Clusterzahl entsprechen; es können z.B. identische Cluster generiert werden, die eine unterschiedliche Entwicklung aufweisen, oder es können zunächst leere Cluster angelegt werden, die sich in der Entstehung befinden und entsprechend erst später Objekte erhalten. Ferner besteht die Möglichkeit, durch

einzelne, fast leere Cluster mit einem sehr hohen Volumen Ausreißer zu simulieren, um ihren Einfluss auf die übrigen Cluster zu messen; diese Cluster sind für die anschließende Clusteranalyse bedeutungslos.

- *Angabe über Clusterform*: Mittels einer Checkbox wird festgelegt, ob es sich um sphärische oder ellipsoide Cluster handelt. Dies ist für spätere Eingabeoptionen entscheidend, da je nach Clusterart unterschiedliche Parameter benötigt werden.
- *Anzahl Dimensionen*: Die Anzahl der Dimensionen ist im Rahmen der dynamischen Untersuchung nicht veränderbar.
- *Anzahl Perioden*: Die Anzahl an Perioden gibt an, wie viele einzelne Datensätze insgesamt generiert werden, d.h., über welche Zeitspanne sich das Experiment erstreckt.
- *Veränderbarkeit der Objekte*: Durch eine Checkbox wird bestimmt, ob es sich um konstante Objekte handelt, die über die Zeit unverändert sind und entsprechend in jeder Periode wieder auftreten, oder ob in jeder Periode eine Anzahl neuer Objekte zufällig generiert werden soll. Der erste Fall ist für eher theoretische Experimente gedacht, in denen nur eine tatsächliche Änderung analysiert werden soll. Für die Experimente bzgl. der allgemeinen Eignung der Ansätze ist der zweite Fall entscheidend, wenn neben den vorgegebenen auch zufällige, durch die Datengenerierung verursachte Änderungen vorliegen, die im Rahmen der Analyse jedoch nicht ins Gewicht fallen dürfen.

Nach Festlegung der allgemeinen Parameter für die Datengenerierung erfolgt die Eingabe der clusterspezifischen Parameter:

- *Objektzahl*: Die Objektzahl je Cluster wird für die erste Periode festgelegt; ferner wird die Änderung der Objektzahl je Cluster und Periode bestimmt. Dabei erfolgt eine Plausibilitätsprüfung, da die Objektzahl eines Clusters im Zeitablauf nicht negativ werden darf.
- *Clusterzentrum*: Für jedes Cluster wird das Zentrum zu Beginn der Analyse festgesetzt. Außerdem erfolgt eine Eingabe der Änderungen je Cluster und Periode in den einzelnen Dimensionen.
- *Kovarianzmatrix*: Analog zu den Clusterzentren erfolgt eine Eingabe der Kovarianzmatrizen je Cluster zu Beginn der Analyse und ihrer Veränderungen je Periode. Dabei unterscheiden sich die Eingabefenster je nach Clusterform:
  - ellipsoide Cluster: Neben den vollständigen Kovarianzmatrizen und ihrer absoluten Änderungen können Rotationen der Daten je Periode über die Kovarianzmatrix realisiert werden. Hierzu erfolgt eine Eingabe der möglichen Drehungen bzgl. verschiedener Dimensionen in Gradangaben für die einzelnen Perioden. Darauf aufbauend wird eine Rotationsmatrix erstellt, die mit der angepassten Kovarianzmatrix aggregiert wird. Dabei ist zu beachten, dass bei der Betrachtung mehrerer aufeinanderfolgender Perioden die Anpassung der Kovarianzmatrizen und die Zusammenfassung mehrerer Rotationen separat erfolgen, bevor die Matrizen mittels Matrizenmultiplikation aggregiert werden.
  - sphärische Cluster: Bei nichtellipsoiden Clustern erfolgt lediglich die Angabe der Varianz der ersten Dimension sowie ihre Änderung je Periode; die übrigen Varianzen werden im Eingabefenster sofort übernommen und sind nicht editierbar. Die Kovarianzen sind konstant auf Null gesetzt. Im Eingabefenster gibt es keine Möglichkeit zur Angabe von Rotationen, da diese bei sphärischen Clustern nicht von Bedeutung sind.

Während der Eingabe der Kovarianzmatrizen erfolgt eine Plausibilitätsprüfung. Varianzen dürfen nicht negativ werden, die Kovarianzen müssen symmetrisch und die resultierenden Korrelationen sinnvoll sein.

Neben der allgemeinen Plausibilitätsprüfung der einzelnen Werte wird bei der Dateneingabe geprüft, ob die benötigten Formate (z.B. double, int) eingehalten werden. Ferner erfolgt eine Anpassung der Werte, so dass Kommata durch Punkte ersetzt und Leerstellen eliminiert werden, damit einfache Eingabefehler keine Fehlerquelle bei der Formatprüfung darstellen.

Analyseparameter

Für den zweiten Funktionsbereich, die Fuzzy-Clusteranalyse im dynamischen Kontext, erfolgt eine separate Dateneingabe der Analyseparameter. Neben allgemeinen Parametern zur Clusteranalyse und dem generellen Aufbau der Experimente werden die spezifischen Parameter gemäß Kapitel 5 für die Analyse dynamischer Veränderungen benötigt.

Folgende allgemeine Analyseparameter sind erforderlich:

- *Clusterzahl für die erste Analyse*: Die Clusterzahl kann u.U. von der Anzahl der Cluster bei der Datengenerierung abweichen, so dass Änderungen wie z.B. eine Clusterteilung sowie das Vorhandensein von Ausreißern simuliert werden können (s.o.).
- *Wahl des Algorithmus*: Durch ein Drop-Down-Menü kann der gewünschte Algorithmus ausgewählt werden. Je nach gewähltem Algorithmus werden weitere notwendige Eingabefelder sichtbar:
  - probabilistische Analyse: Für die probabilistische Analyse (Abschnitt 3.3) sind keine weiteren Eingabefelder erforderlich. Ihre Aufnahme in das Applet erfolgte lediglich zu Vergleichszwecken, eine generelle Anwendung im dynamischen Kontext ist nicht sinnvoll (vgl. Abschnitt 3.4 zur Bedeutung der possibilistischen Clusteranalyse).
  - klassische possibilistische Analyse: Für die klassische possibilistische Analyse nach Krishnapuram und Keller (1993) (Abschnitt 3.4) wird zusätzlich eine Checkbox benötigt, mittels der festgelegt wird, ob eine einmalige Neuberechnung der η_i erfolgen soll. Auch die klassische possibilistische Analyse wurde lediglich zu Vergleichszwecken in das Applet integriert.
  - Heterogenitätsansätze zur Modellierung der Clusterabstoßung gemäß Abschnitt 4.2: Neben der Festlegung zur Neuberechnung der η_i erfolgt eine Eingabe des Gewichtungsparameters γ für den Bestrafungsterm sowie des allgemeinen Normalisierungsparameters ζ für die Prototypdistanzen. Diese Ansätze sind jedoch nur zum Nachverfolgen einiger bestimmter Änderungen wie das Aufdecken neu entstehender und gefährderter Cluster geeignet (vgl. Abschnitt 4.2).
  - Homogenitätsansätze gemäß Abschnitt 4.3: Zu den für die Heterogenitätsansätze benötigten Parametern werden die Variante zur Bestimmung des clusterspezifischen Normierungsparameters ζ_i festgesetzt und der Grenzwert α für den α-Schnitt bestimmt.
- *Parameter für die dynamische Analyse*: Für die dynamische Analyse werden verschiedene allgemeingültige Parameter benötigt:

  - Länge des Zeitfensters τ: Durch die Zeitfensterlänge τ wird die Anzahl der Perioden bestimmt, die jeweils in eine Analyse einbezogen werden. Auf diese Weise wird auch der erstmögliche Analysezeitpunkt festgesetzt.
  - Analysefrequenz Δt: Die Analysefrequenz legt fest, in welchen Abständen, d.h. nach wie vielen Perioden, jeweils eine erneute Analyse stattfinden soll.
  - Grenzwert der Absorption α^{A}: Für die Zuordnung neuer Objekte zu bekannten Clustern wird der Absorptionsgrenzwert bestimmt.
  - Gewichtung neuer Daten bei inkrementellen Updates: Für die Gewichtung neuer Daten im Vergleich zu den älteren, in den vorhandenen Clusterprototypen berücksichtigten Daten werden die Parameter $\gamma^{\vec{v}}$ bzw. γ^{Σ} festgesetzt.
- *Festsetzung, ab wann eine Änderung als signifikant anzusehen ist*: Werden Änderungen gradueller Art aufgedeckt, wird ein Parameter $\beta^{\vec{v}}$ benötigt, der festlegt, inwiefern diese Änderungen adaptiert oder als nicht signifikant verworfen werden (vgl. Abschnitt 5.3).
- *Anzahl Analysen*: Zusätzlich kann angegeben werden, wie viele Analysen mit den eingegebenen Parametern durchgeführt werden können. Durch die wiederholte Durchführung eines Experiments mit vorgegebenen Parametern kann überprüft werden, ob Ergebnisse eher zufällig entstanden sind oder allgemeine Gültigkeit besitzen.

Neben den allgemeinen Analyseparametern werden Parameter für die Untersuchung abrupter Veränderungen innerhalb einer Clusterstruktur benötigt. Folgende Parameter werden hierzu erhoben:

- *Neubildung von Clustern*:
  - $\beta_{nC}^{\min}$ zur Bestimmung der geforderten Mindestgröße und -dichte eines potentiell entstehenden Clusters
  - β_{nC} zur Bestimmung der geforderten Mindestgröße und -dichte eines bereits entstandenen Clusters
  - $\delta_{\min}$ minimales Wachstum eines Ausreißerclusters bei Nachverfolgung der Entwicklung, um weiterhin als potentielles Cluster angesehen zu werden
  - $t_{nC}^{\max}$ als Zeitdauer, über die der Grenzwert $\delta_{\min}$ bei der Entstehung eines neuen Clusters maximal unterschritten werden darf
- *Clusterelimination*:
  - $\beta_{eC}^{\min}$ zur Bestimmung der geforderten Mindestgröße und -dichte eines noch existierenden, gefährdeten Clusters
  - β_{eC} zur Bestimmung der geforderten Mindestgröße und -dichte eines ungefährdeten Clusters
  - $\lambda_{\min}^{neu}$ als Grenzwert für die minimale Neuzuordnung von Objekten je Periode: Neben der Eingabe eines Grenzwerts erfolgt anhand von Radiobuttons eine Auswahl, ob es sich um einen prozentualen Wert handelt, die Wahl also im Vergleich zur Größe der übrigen Cluster erfolgt, oder ob es sich um einen kontextabhängig festgesetzten absoluten Wert handelt.
  - $t_{eC}^{\max}$ als Zeitdauer, über die der Grenzwert $\lambda_{\min}^{neu}$ maximal unterschritten werden darf, bevor ein Cluster eliminiert wird

- *Clustervereinigung*:
  - γ^s zur Gewichtung der Distanz im Vergleich zur Parallelität bei der Aggregation der Werte zur Bestimmung der Ähnlichkeit zweier Cluster
  - $\lambda_V^{\mathrm{I}^{\max}}$ als Grenzwert für den Überschneidungsgrad, von dem an eine Vereinigung zweier Cluster ohne weitere Prüfung der Distanz- und Parallelitätswerte erfolgt
  - $\lambda_V^{\mathrm{I}^{\min}}$ als Grenzwert für den minimalen Überschneidungsgrad, von dem an für eine Vereinigung zweier Cluster eine weitere Prüfung der Distanz- und Parallelitätswerte durchgeführt wird
  - $\lambda_{\max}^s$ als Grenzwert für die Ähnlichkeit zwischen zwei Clustern nach Aggregation ihrer Distanz- und Parallelitätswerte, von dem an eine Vereinigung zweier Cluster erfolgt
  - $\lambda_{\min}^s$ als Grenzwert für die minimale Ähnlichkeit zwischen zwei Clustern nach Aggregation ihrer Distanz- und Parallelitätswerte, von dem an zwei Cluster für eine mögliche zukünftige Vereinigung vorgemerkt werden
- *Clustertrennung*:
  - $\alpha_\kappa^{\mathrm{A}}$ als Grenzwert der Absorbierung zur Differenzierung zwischen einer möglichen Clusterelimination und einer -trennung auf Basis des Kompaktheitsindexes
  - α_{SC}^{A} als Grenzwert der Absorbierung für die Extraktion der zum Subclustering innerhalb eines Clusters verwendeten Objekte
  - λ_{par}^{CT} als Grenzwert für die minimal benötigte Clusterdrehung bei der Untersuchung einer möglichen Clustertrennung
  - λ_λ^{CT} als Grenzwert für die minimal geforderte Ausrichtungsänderung für eine mögliche Clustertrennung
  - λ_{dist}^{CT} als Grenzwert für die minimal benötigte Distanz zwischen zwei Subclusterzentren, um als geteilt gelten zu können
  - $\lambda_{CT}^{\mathrm{I}^{\min}}$ als Grenzwert für die maximal zulässige Überschneidung zweier Subcluster, um als bereits geteilt zu gelten
  - $\lambda_{CT}^{\mathrm{I}^{\max}}$ als Grenzwert für die maximal zulässige Überschneidung zweier Subcluster, um als potentiell teilbar zu gelten
  - $\beta_{CT}^{\min}$ zur Bestimmung der geforderten Mindestgröße und -dichte eines Subclusters für eine potentielle Clustertrennung
  - β_{CT} zur Bestimmung der geforderten Mindestgröße und -dichte eines Subclusters für eine vollzogene Clustertrennung

Für die einzelnen Analyseparameter erfolgt analog zur Eingabe der für die Datengenerierung benötigten Daten eine Format- und Plausibilitätsprüfung, so dass parameterabhängig definierte Intervalle in jedem Fall eingehalten werden müssen.

Alle Werte werden sofort gespeichert, damit bei einem Zurück- und Vorblättern zwischen den einzelnen Eingabefenstern keine erneute Eingabe bereits vorhandener Daten erfolgen muss. Dasselbe gilt, wenn im Anschluss an ein abgeschlossenes Experiment ein neues Experiment durchgeführt werden soll: Auch hier werden zur Vereinfachung der Dateneingabe die Werte des vorherigen Experiments übernommmen.

C.1.2. Durchführung der Analyse

Nach vollständiger Eingabe aller Daten erfolgt die eigentliche Datenanalyse, die je Periode aus mehreren Teilschritten besteht:

1. *Datengenerierung*: Für die Experimente werden in jeder Periode standardnormalverteilte (Pseudo-)Zufallsdaten mit Hilfe der `Random`-Klasse und der zugehörigen `nextGaussian()`-Methode generiert, die im Anschluss unter Berücksichtigung der eingegebenen Clusterparameter entsprechend transformiert werden[25].
2. *Evaluation vorhandener Veränderungen*: Zu Analysezeitpunkten, die durch die Zeitfensterlänge τ und die Analysefrequenz Δt festgesetzt sind, erfolgt eine Evaluation der aufgetretenen graduellen Änderungen unter Einbeziehung der inkrementellen Updates der Clusterzentren sowie – bei ellipsoiden Clustern – der Kovarianzmatrizen und eine Bestimmung der lokalen Parameter. Ferner werden abrupte Veränderungen gemäß den Abschnitten 5.4 bis 5.7 untersucht und nachverfolgt, außerdem wird die aktuelle, für die Analyse benötigte Clusterzahl bestimmt. Die Evaluation der vorhandenen Veränderungen erfolgt erstmals in Periode $\tau + \Delta t$, da vorher noch kein Vergleich möglich ist.
3. *Durchführung der Fuzzy-Clusteranalyse*: Zu allen Analysezeitpunkten, d.h. beginnend in Periode τ, wird eine Clusteranalyse anhand der vorhandenen, ggf. im Rahmen der Evaluation der Veränderungen angepassten Clusterzahl durchgeführt. Zu späteren Analysezeitpunkten werden zur Initialisierung die Ergebnisse vorangegangener Untersuchungen verwendet.

C.1.3. Ausgabefenster

Vor der Ausgabe der analysebezogenen Ergebnisse erfolgt zunächst eine Ausgabe der für ein Experiment vorgegebenen Parameter, so dass gespeicherte Ergebnisse auch später nachvollzogen werden können. Im Anschluss werden die Durchführungsschritte und Ergebnisse je Periode einzeln aufgeführt; in Perioden, in denen keine Analyse erfolgt, wird zur besseren Übersicht ein Hinweis auf die nicht durchgeführte Analyse gegeben. Zu den einzelnen Analysezeitpunkten werden die aktuellen, in die Analyse einbezogenen Daten ausgegeben und die Analyseergebnisse hinzugefügt. Dabei erfolgt eine mehrstufige Ausgabe einzelner Teilergebnisse:

1. *Änderung der Parameter*: Zunächst erfolgt die Ausgabe der Änderung der Clusterzentren, der harten Kovarianzmatrizen sowie der verschiedenen lokalen Maße gemäß Abschnitt 5.3. Dabei wird zwischen „echten“ Veränderungen, d.h. in den generierten Daten mit bekannter Clusterzugehörigkeit vorhandenen Schwankungen, und den durch die Analyseergebnisse implizierten Änderungen unterschieden, um eine Nachprüfung der aufgedeckten Änderungen zu ermöglichen.
2. *Ausgabe potentiell entstehender Cluster*: Werden im Rahmen der Analyse Ausreißercluster aufgedeckt, werden jeweils Größe, Dichte und Fuzzy-Kardinalität inklusive der zugehörigen Minimalwerte sowie die Prototypen ausgegeben. Ferner wird angegeben, ob sich ein

[25] Bei nicht veränderbaren Objekten erfolgt i.d.R. anstelle einer Neugenerierung der Daten lediglich eine Übernahme der vorhandenen Datensätze unter Berücksichtigung der gegebenen Objektzahl. Nur wenn mehr Objekte als vorhanden benötigt werden, werden diese entsprechend der beschriebenen Vorgehensweise generiert.

Cluster noch in der Entstehung befindet oder bereits als vollwertig entstandenes Cluster anzusehen ist. Sind aus einer vorangegangen Analyse potentielle Cluster vorhanden, folgen eine separate Prüfung dieser Cluster und Angaben zur ihrem Entwicklungsstand.

3. *Ausgabe gefährdeter Cluster*: Gilt ein Cluster aufgrund der Analyse zur Clusterelimination als gefährdet, wird diese Entwicklung unter Angabe der einzelnen Parameter und der Vergleichswerte dokumentiert. Entsprechend wird ein zu eliminierendes Cluster ausgegeben. Die Nachverfolgung der in vorangegangenen Analysen als gefährdet markierter Cluster erfolgt ebenfalls unter Ausgabe des Entwicklungsstands bzgl. Größe und Dichte. Liegen weder gefährdete noch zu eliminierende Cluster vor, wird dieses durch einen Hinweis angemerkt.
4. *Ausgabe potentiell zu vereinigender Cluster*: Werden mögliche zu vereinigende Cluster erkannt, erfolgt eine Ausgabe der entsprechenden Parameter bzgl. ihres Überschneidungsgrads auf Basis des Inklusionsmaßes, ihrer Ähnlichkeit nach Aggregation der Distanz- und Parallelitätswerte sowie der zugehörigen Vergleichswerte, die für die Bestimmung der Distanz- und Parallelitätsfunktionen hinzugezogen wurden. Außerdem wird dokumentiert, ob es sich um eine zukünftig mögliche Clustervereinigung handelt oder ob diese bereits stattgefunden hat. Analog zu den bereits aufgeführten abrupten Veränderungen erfolgt ferner eine Nachverfolgung potentiell zu vereinigender Cluster aus den vorangegangenen Analysen. Ist keine Clustervereinigung innerhalb der Clusterstruktur möglich, wird auf diesen Umstand hingewiesen.
5. *Ausgabe potentiell zu trennender Cluster*: Wird für ein Cluster eine mögliche Teilung aufgedeckt, werden die vorhandenen Parameter zur Änderung der clusterinternen Struktur ausgegeben. Bei ausreichender Änderung der clustereigenen Strukturschwankungen wird ein Subclustering durchgeführt und die resultierenden Prototypen der einzelnen Subcluster werden dokumentiert. Außerdem erfolgt eine Ausgabe der zur Bestimmung einer bereits stattgefunden Trennung benötigten Parameter bzgl. Distanz der Subclusterzentren, Überlappungsgrad der Subcluster und Größe und Dichte der einzelnen Subcluster sowie der notwendigen Vergleichswerte. Ferner erfolgt eine Nachverfolgung potentiell zu trennender Cluster aus den vorangegangenen Analysen.
6. *Ausgabe der Ergebnisse der Fuzzy-Clusteranalyse*: Bei Durchführung einer Fuzzy- Clusteranalyse werden die Ergebnisse zu den einzelnen, in der Analyse aufgedeckten Clusterprototypen ausgegeben.

C.2. Durchführung der Experimente

Zur Untersuchung der Eignung der einzelnen in Kapitel 5 eingeführten Maße wurde eine Vielzahl an Experimenten durchgeführt, die bzgl. ihrer Parameter zum Teil stark variierten. Neben den stark vereinfachten Beispielen, die zur besseren Nachvollziehbarkeit in Kapitel 5 ausführlich dokumentiert sind, erfolgten weitere, oftmals komplexere Betrachtungen der Veränderungen, so dass die Gültigkeit der Ergebnisse auch für andere Parameterwerte evaluiert werden konnte. Im Wesentlichen unterschieden sich die Experimente bzgl. folgender Größen:

- *Anzahl Dimensionen*: Neben grafisch darstellbaren Datensätzen im zwei- bis dreidimensionalen Raum wurden Experimente bis hin zum $\mathbb{R}^5$ betrachtet. Da in höherdimensionierten Räumen keine grafische Darstellung mehr möglich ist, wurde in diesen Fällen auf eine Ausgabe der generierten Objektdaten verzichtet.

- *Anzahl Perioden*: Abhängig von den Experimenten wurden unterschiedliche Periodenzahlen festgelegt, i.d.R. wurden insgesamt 5-15 Perioden betrachtet. Gerade bei langsam verlaufenden Änderungen ist die Betrachtung eines längeren Zeitraums erforderlich, um die Änderungen nachvollziehen zu können. Dies gilt insbesondere bei einer Clustertrennung: Ausgehend von identischen Clusterprototypen, die sich in unterschiedliche Richtungen entwickeln, dauert es – je nach Geschwindigkeit der Änderung – oftmals lange, bis überhaupt eine Änderung erkennbar wird, da sich die Cluster in den ersten Perioden noch derartig stark überlappen, dass keine gesonderte Entwicklung aufgedeckt werden kann (vgl. Abschnitt 5.7).
- *Anzahl Cluster*: Die zur Datengenerierung verwendete Clusterzahl variierte aufgrund der Simulation von Ausreißern sehr stark; teilweise wurden sehr viele Cluster mit einer hohen Ausdehnung innerhalb der einzelnen Dimensionen kreiert, für die jeweils nur eine vernachlässigbar geringe Anzahl an Objekten generiert wurde. Die Anzahl „echter“, d.h. für die Analyse relevanter Cluster beschränkte sich in den durchgeführten Experimenten auf maximal sieben, so dass verschiedene Entwicklungen gleichzeitig simuliert werden konnten.
- *Anzahl Objekte*: In den einzelnen Experimenten wurden deutlich unterschiedliche Objektzahlen verwendet, sowohl insgesamt als auch pro Cluster. Einige Anpassungen bzgl. der Objektzahl je Cluster ergeben sich bereits aus den simulierten Änderungsarten: so startet z.B. ein entstehendes Cluster i.d.R. ohne Objekte, die Objektzahl wird im Anschluss periodenweise sukzessiv erhöht. Außerdem wurden unterschiedliche Strukturen betrachtet, d.h. zum einen Cluster, die – sofern vollständig vorhanden – etwa dieselbe Objektzahl vorweisen, zum anderen sehr unterschiedlich große Cluster, um festzustellen, ob auch Änderungen in kleineren Clustern nachvollzogen werden können, wenn große Cluster innerhalb der allgemeinen Clusterstruktur dominieren. Ferner schwankte die Anzahl an generierten Ausreißern, um abhängig von der allgemeinen Objektzahl innerhalb der generierten „echten“ Cluster zu testen, ab welchem Zeitpunkt keine geeignete Clusterstruktur mehr aufgedeckt werden kann. Letztere Untersuchung war insbesondere für die Beurteilung der einzelnen Clusteranalyseverfahren von Bedeutung. Die probabilistische Analyse (Abschnitt 3.3) zeigt eine hohe Abhängigkeit von der Anzahl der Ausreißer, ebenso die klassische possibilistische Analyse nach Krishnapuram und Keller (1993) (Abschnitt 3.4). Die betrachteten Erweiterungen der possibilistischen Analyse (Kapitel 4) sind hingegen weniger anfällig gegenüber Ausreißern, da sie stärker auf die einzelnen Cluster fokussieren.
- *Clusterstruktur*: Die erstellten Cluster variierten stark bzgl. ihrer Ausdehnung, ihrer Lage und ihrer generellen Struktur. So wurden analog zur Anpassung der Objektzahlen auch bzgl. des Clustervolumens zum einen Datensätze mit eher gleichartigen Clustern kreiert, zum anderen wurden Clusterstrukturen erzeugt, bei denen sich die einzelnen Cluster bzgl. ihres Volumens und ihrer Dichte stark unterschieden. Hierbei war sehr auffällig, dass die Ansätze zur klassischen possibilistischen Analyse nach Krishnapuram und Keller (1993) (Abschnitt 3.4) bei sehr unterschiedlichen clusterinternen Strukturen kaum in der Lage waren, Cluster von geringerer Dichte aufzudecken. Des Weiteren schwankte die allgemeine Clusterstruktur dahingehend, dass neben dem Vorkommen von Ausreißern auch die Separierung der einzelnen Cluster untereinander stark voneinander abwich, so dass eine Beurteilung bzgl. der Eignung einzelner Parameter bei eher undeutlichen Clusterstrukturen möglich war. Zusammenfassend lässt sich festhalten, dass die Veränderungen umso deutlicher hervortraten, je eindeutiger die Cluster separiert sind, da der Einfluss

der übrigen Cluster geringer ist.

- *Zeitfensterlänge*: Für die Zeitfensterlänge τ wurden verschiedene Werte getestet. Dabei ergab sich, dass kleinere Werte für τ zu bevorzugen sind, damit der Anteil neuerer Daten in den einzelnen Analysen nicht zu gering ist.
- *Analysefrequenz*: Für die Analysefrequenz Δt wurden ebenfalls unterschiedliche Werte getestet. Hierbei zeigte sich, dass häufige Analysen erforderlich sind, um Änderungen gezielt aufdecken und prognostizieren zu können. Bei langen Zeitfenstern zwischen den einzelnen Analysen erfolgen die Veränderungen teilweise derartig stark, dass die auftretenden Veränderungen nur noch bedingt nachvollzogen, geschweige denn vorhergesagt werden können.
- *Parameter zur Untersuchung der Veränderungen*: Für die einzelnen Parameter, die zur Analyse im dynamischen Kontext benötigt werden, wurden verschiedene Werte in den vorgegebenen Intervallen getestet, analog zu den Experimenten für eine geeignete Wahl von α beim α-Schnitt in den Ansätzen zur possibilistischen Analyse unter Einbeziehung der Clusterhomogenität (vgl. Abschnitt 4.4.2). Generell ist die Wahl der einzelnen Parameter dabei stark struktur- und kontextabhängig. Je eindeutiger die allgemeine Clusterstruktur separiert ist und je deutlicher die Veränderungen simuliert wurden, desto leichter ließen sich Änderungen nachvollziehen. Ab wann eine Änderung als signifikant bzw. als vollzogen anzusehen ist, hängt von den Rahmenbedingungen und dem Ziel der einzelnen Untersuchungen ab.
- *Änderungen innerhalb der Clusterstruktur*: Neben einfachen Experimenten, in denen nur einzelne Veränderungen innerhalb der Clusterstruktur erfolgten, wurden außerdem verschiedene Änderungen gleichzeitig durchgeführt, d.h., es wurden sowohl verschiedene Änderungen innerhalb eines Clusters simultan vorgenommen als auch unterschiedliche Cluster zeitlich parallel angepasst. Je komplexer die Änderungen und je weniger eindeutig die Clusterstruktur insgesamt waren, desto häufiger traten Messfehler auf; diese hielten sich jedoch soweit im Rahmen, dass eine allgemeine Untersuchung abrupter Veränderungen weiterhin im Bereich des Möglichen lag. Werden Änderungen allerdings zu schnell durchgeführt, z.B. durch ein sehr schnelles Auseinanderdriften eines einzelnen Clusters, kann es vorkommen, dass anstelle der tatsächlichen Entwicklung des einzelnen Clusters andere Änderungen als Ursache angenommen werden. Im Beispiel wird teilweise das Ursprungscluster eliminert, während zwei neue aufgedeckt werden, oder aber es wird aufgrund des beschränkten Betrachtungsraums ein Clusterdrift eines Clusters festgestellt, während an anderer Stelle ein neu entstehendes Cluster angenommen wird.

Danksagungen

An der Erstellung dieser Arbeit waren indirekt mehrere Personen beteiligt. Um es mit den Worten Homunkoloss' aus Walter Moers' „*Die Stadt der träumenden Bücher*“ zu sagen:

> „Es ist wie beim Schreiben eines Romans: Am Anfang ist alles ganz leicht, die ersten Kapitel schreiben sich mit dem größten Schwung. Aber dann wirst du irgendwann müde, du blickst zurück und siehst, dass du erst die Hälfte hinter dir hast. Du blickst nach vorn und siehst, dass die andere Hälfte noch vor dir liegt. Wenn du dann den Mut verlierst, bist du verloren. Es ist leicht, etwas zu beginnen. Es ist schwer, etwas zu Ende zu bringen.“

Ohne die Unterstützung anderer ist es kaum möglich, eine solche Arbeit zu erstellen, ohne die Motivation und den Glauben an das Gelingen der Arbeit zu verlieren. Den Menschen, die mich dabei begleitet und vorangebracht haben, möchte ich an dieser Stelle danken.

Mein Dank gilt meinem Doktorvater Herrn Professor Dr. K. Ambrosi. Das Vertrauen, das Sie mir entgegengebracht haben, hat mich darin bestärkt, dass ich meine Dissertation erfolgreich zu Ende bringen kann. Ohne Ihre Ideen und Anregungen wäre diese Arbeit nicht möglich gewesen.

Für die Übernahme des Zweitgutachtens bin ich Herrn Professor Dr. D. Baier zu Dank verpflichtet. Danke für Ihren besonderen Einsatz, durch den es möglich war, die terminliche Dringlichkeit meinerseits zu berücksichtigen. Ebenso danke ich den übrigen Mitgliedern der Promotionskommission für die Zeit, die Sie in die Auseinandersetzung mit meiner Arbeit investiert haben.

Meinen Kollegen danke ich für ihre Unterstützung durch Diskussionen und Ideen, die mich voranbrachten, mir neue Wege aufzeigten und mir so eine große Hilfe waren.

Die Promotion war außerdem nur möglich durch die Hilfe, die mir aus dem Kreis meiner Familie entgegengebracht wurde. Besonderer Dank gilt dabei meinen Eltern. Danke nicht nur für all die Unterstützung, die ihr mir während meines Studiums habt zuteil werden lassen. Vielen, vielen Dank insbesondere dafür, dass ihr mir auch die Möglichkeit zur Promotion gegeben habt, indem ihr euch stets so liebevoll um Lara gekümmert habt. Meiner Mutter danke ich außerdem für die Lektorentätigkeit bei meiner Dissertation. Auch meine Schwiegermutter ist hervorzuheben. Danke, dass du gerade in der Endphase meiner Arbeit regelmäßig deine Sonntage für Laras Betreuung geopfert hast, damit ich arbeiten konnte.

Seit Februar 2011 ist unsere Tochter Lara ein wichtiger Teil unseres Lebens. Auch wenn du noch nicht begreifst, was das alles bedeutet: Danke dir, dass du mir Antrieb gegeben hast, meine Arbeit fertigzustellen, dass du mich durch deine fröhliche Art immer wieder aufgemuntert hast, wenn ich niedergeschlagen war. Und entschuldige bitte all die Male, in denen ich gesagt habe, dass ich keine Zeit zum Spielen hätte, weil ich arbeiten müsste. Wir können aber auch nach Abschluss der Arbeit gerne weiterhin Karusell auf meinem Schreibtischstuhl fahren.

Den wichtigsten Part spielte jedoch mein Ehemann Dr. Marcel Minke. Ich kann kaum in Worte fassen, wie dankbar ich dir bin, dass du all meine Launen ertragen hast, dass du mir mit Rat und Tat zur Seite gestanden hast, dass du mich aufgebaut hast, wenn ich nicht mehr weiter

wusste, dass du mich immer wieder angetrieben und mich mit so viel Verständnis und Liebe unterstützt hast. Du bist mein Ruhepol. Danke, dass du da bist und mich immer unterstützt, was ich auch anfange.

Zum Schluss möchte ich noch einen ganz kleinen Dank aussprechen, und zwar unserem zweiten Kind. Es ist noch gar nicht geboren, doch es ist der Grund, aus dem diese Arbeit jetzt fertig ist. Durch dieses Kind habe ich den Druck bekommen, den ich brauchte, um meine Dissertation abzuschließen.

Glossar

Attribut
→ Eigenschaft

Ausreißer
Objekt innerhalb einer bekannten Clusterstruktur, das keinem Cluster zugeordnet ist.

Ausreißercluster
Dichte Region innerhalb von Ausreißern, die Entstehung eines neuen Clusters impliziert.

Change Mining
Change Mining beschreibt Data Mining-Ansätze zur Analyse der dynamischen Entwicklung von Mustern.

Clustertrajektorien
Bewegungspfad eines Clusterprototyps, der die Veränderungen innerhalb der Clusterhistorie enthält.

Data Mining
Das Anwenden eines spezifischen Algorithmus im Rahmen des KDD-Prozesses wird als Data Mining bezeichnet.

Eigenschaft
beschreibende Charakteristik eines Untersuchungsgegenstandes; auch Attribut, Merkmal

Knowledge Discovery in Databases
Knowledge Discovery in Databases (KDD) beschreibt den Prozess der Datenanalyse, um in großen Datenbanken nach Mustern zu suchen und vorgegebene Strukturen zu prüfen.

Merkmal
→ Eigenschaft

Muster
In den Daten vorhandene Zusammenhänge zwischen den Objekten, die durch die Bestimmung eines spezifizierten Modells im Rahmen des Data Mining aufgedeckt werden sollen, werden auch als Muster oder Struktur bezeichnet.

Objekt
Zu analysierender Untersuchungsgegenstand, der durch seine Eigenschaftsausprägungen im Objektvektor spezifiziert wird.

Periode
Erhebungszeitraum, in dem Daten gesammelt werden (z.B. Woche, Monat, Quartal)

Reclustering
Erneute Clusteranalyse, bei der zur Initialisierung zuvor bestimmte Clusterprototypen verwendet werden.

Struktur
→ Muster

Subcluster
Teilcluster, das bei Trennung eines Gesamtclusters entsteht

Subclustering
Clusteranalyse zur Zerlegung eines Gesamtclusters in mehrere Teil- bzw. Subcluster

Zeitfenster
für eine Analyse herangezogene Anzahl an Perioden; auch: Zeitintervall

Zeitintervall
→ Zeitfenster

Abkürzungsverzeichnis

INDSCAL	Individual Differences Scaling
KDD	Knowledge Discovery in Databases
MDS	Multidimensionale Skalierung
RMT	Rule Matching Threshold

Symbolverzeichnis

$a\,(X_1, \ldots, X_T)$	Anpassungsfunktion bei der simultanen Dynamisierung (MDS)
A	Itemmenge (Association Rule Mining)
A_i	Positiv definite Normmatrix der Kovarianzmatrix, $A_i = \det(\Sigma_i)^{\frac{1}{p}} \Sigma_i^{-1}$
$APD\,(U, C)$	Mittlere Partitionsdichte zur Beurteilung der Clustervalidität
$b\,(\vec{x}_1, \ldots, \vec{x}_n)$	Rohstress: Gütekriterium der nichtmetrischen MDS
$b_{\text{INDSCAL}}\,(X_1, \ldots, X_G)$	Stressfunktion bei Individual Differences Scaling (INDSCAL)
$b_{\max}$	maximal möglicher Stresswert bei der nichtmetrischen MDS
$b_{sim}\,(D_1, \ldots, D_T, X_1, \ldots, X_T)$	Stressfunktion bei der simultanen Dynamisierung (MDS)
B	Itemmenge (Association Rule Mining)
c	Clusterzahl (bzw. Anzahl relevanter Assoziationsregeln)
c_t	Anzahl Cluster in Periode t
c_t^{pot}	Maximale Anzahl potentiell neuer Cluster in Periode t
C	Menge der Clusterprototypen; bei c Clustern: $C = \{C_1, \ldots, C_c\}$
C_i	Prototyp eines Clusters i
$C_{t_i}^{\text{A}}$	i-tes Ausreißercluster zum Zeitpunkt t, enthält entstehungsrelevante Parameter
$\text{card}(Y_i)$	Fuzzy-Kardinalität der Menge Y_i, d.h. $\text{card}(Y_i) = \sum_{j=1}^{n} u_{ij}$
d_{jk}^{g}	erhobene Distanz zwischen den Objekten j und k in g-ter Konfiguration (MDS)
$d_{jk}^{tt'}$	erhobene Distanz zwischen Objekt j zum Zeitpunkt t und Objekt k zum Zeitpunkt $t^{'}$
$\hat{d}\,(\vec{x}_j, \vec{x}_k)$	Proximität bzgl. der Objektvektoren $\vec{x}_j$ und $\vec{x}_k$ in einer Konfiguration (MDS)
$d^2\,(C_i, C_{i'})$	quadrierte Distanz zwischen Clusterprototypen C_i und $C_{i'}$
$d_{C_i}^2\,(\vec{v}_i, \vec{x}_j)$	quadrierte Distanz zwischen Clusterzentrum $\vec{v}_i$ und Objekt j mit Eigenschaftsvektor $\vec{x}_j$ unter Berücksichtigung der Eigenschaften des Clusterprototypen C_i
$d_{C_i}^2\,(\vec{v}_i, \vec{x}_j, \vec{x}_k)$	Kombination der Distanzen bzgl. der Objektvektoren $\vec{x}_j$ und $\vec{x}_k$ sowie dem Clusterzentrum $\vec{v}_i$ unter Berücksichtigung der Eigenschaften des Clusterprototypen C_i

D	$c \times n$-Matrix der Distanzen zwischen Clusterzentren $\vec{v}_i$ und Objekten $\vec{x}_j$
D^S	$T \times T$-Superdistanzmatrix, enthält alle erhobenen Distanzmatrizen $D_{tt'}$
D_t	$n \times n$-Matrix der erhobenen Objektdistanzen in Periode t, $t = 1, \ldots, T$
$D_{tt'}$	$n \times n$-Matrix der erhobenen Distanzen zwischen Objekten der Perioden t und t'
$\vec{e}_{il}$	Eigenvektor zum Eigenwert θ_{il} der Kovarianzmatrix von Cluster C_i, $l = 1, \cdots, p$
E	Einheitsmatrix
$f.\left(d^2\left(C_i, C_{i'}\right)\right)$	Funktion zur Clusterabstoßung bzgl. der Clusterprototypen C_i und $C_{i'}$
$F\left(O\right)$	Menge aller Fuzzy-Mengen von O
$FHV\left(\left[U_i\right]_{\alpha^{\mathrm{A}}}\right)$	Fuzzy-Hypervolumen eines Clusters mit Zugehörigkeitsmenge $\left[U_i\right]_{\alpha^{\mathrm{A}}}$
g	Index für Konfigurationen, $g = 1, \ldots, G$
G	Anzahl individueller Konfigurationen (MDS)
$h_\phi\left(D_1, \ldots, D_T, X_1, \ldots, X_T\right)$	Optimierungsfunktion bei der simultanen Dynamisierung (MDS)
$\hbar_{ii'}$	Grad der Übereinstimmung der Folgerungen zweier Assoziationsregeln r_i und $r_{i'}$
i	Index für resultierende Cluster (bzw. Assoziationsregeln), $i = 1, \ldots, c$
I	Idealprodukt
$\mathtt{I}_{ii'}$	Inklusionsmaß zur Bestimmung des Überschneidungsgrades zweier Cluster C_i und $C_{i'}$
$\mathfrak{I}$	Menge aller Items (Association Rule Mining)
j	Index für Objekte, $j = 1, \ldots, n$
$J\left(X, U, C\right)$	Optimierungsfunktion beim Fuzzy-Clustering
k	Index für Objekte, $k = 1, \ldots, n$
$k^1_{ii'}$	quadrierte euklidische Distanz zwischen den Clusterzentren $\vec{v}_i$ und $\vec{v}_{i'}$ (unabhängig von verwendetem Algorithmus)
$k^2_{ii'}$	Skalarprodukt der kleinsten Eigenvektoren zweier Clusterprototypen C_i und $C_{i'}$
K	Konstante
l	Index für Dimensionen, $l = 1, \ldots, p$

$\ell_{ii'}$	Grad der Übereinstimmung der Bedingungen zweier Assoziationsregeln r_i und $r_{i'}$
$L(c_t)$	Fuzzy-Variante der Innergruppenvarianz zu einer Clusterzahl c_t, die als Schadensfunktion zur Bestimmung der Strukturstärke $St(c_t)$ herangezogen wird
$LPC([U_i]_{\alpha^{\mathrm{A}}})$	Lokaler Partitionskoeffizient eines einzelnen Clusters mit Zugehörigkeitsmenge $[U_i]_{\alpha^{\mathrm{A}}}$
m	Fuzzifier zur Steuerung der Unschärfe, $m \in (1, \infty)$; i.d.R. $m = 2$
M	Produktmerkmal
n	Objektzahl (bzw. Anzahl Transaktionen beim Association Rule Mining)
n^{A}	Anzahl Ausreißer, d.h. Objekte, für die gilt $u_{ij} < \alpha^{\mathrm{A}} \forall i = 1, \ldots, c$
n_i^{α}	Anzahl der Objekte, die durch Cluster C_i mit Grenzwert α absorbiert werden
$n_{i_{neu}}^{\alpha^{\mathrm{A}}}$	Anzahl neu von Cluster C_i absorbierter Objekte (seit letztem Analysezeitpunkt)
$n_{i_{pot}}^{\alpha^{\mathrm{A}}}$	Anzahl durch ein potentielles Cluster C_i absorbierter Objekte
$n_{\min}^{\alpha^{\mathrm{A}}}$	Anzahl an Objekten, die durch den α-Schnitt mit α^{A} durch das kleinste (ungefährdete) Cluster absorbiert werden, d.h. $n_{\min}^{\alpha^{\mathrm{A}}} = \left\{ n_i^{\alpha^{\mathrm{A}}} \mid i \in \{1, \ldots, c\} \right\}$
$\Delta n_{t_i}^{\alpha^{\mathrm{A}}}$	Veränderung der Clustergröße eines Ausreißerclusters $C_{t_i}^{\mathrm{A}}$ gegenüber der Vorperiode; falls $t = t_i^0$: $\Delta n_{t_i}^{\alpha^{\mathrm{A}}} := n_{t_i}^{\alpha^{\mathrm{A}}}$
N_i	Normalverteilung von Cluster C_i nach Gath und Geva (1989): $N_i(\vec{v}_i, \Sigma_i)$
O	Objektmenge
p	Dimensionszahl
P	Wahrscheinlichkeit
P_i	Apriori-Wahrscheinlichkeit der Clusterzuordnung zu Cluster C_i nach Gath und Geva (1989)
$PC(U)$	Partitionskoeffizient zur Beurteilung der Clustervalidität
PD_i	Partitionsdichte eines einzelnen Clusters als lokales Maß basierend auf der mittleren Partitionsdichte $APD(U, C)$
$PD_{i_{pot}}$	Dichte eines potentiellen Clusters i
$PD_{\min}$	minimale Dichte aller ungefährdeten Cluster
ΔPD_{t_i}	Veränderung der lokalen Partitionsdichte eines Ausreißerclusters $C_{t_i}^{\mathrm{A}}$ gegenüber der Vorperiode; falls $t = t_i^0$: $\Delta PD_{t_i} := PD_{t_i}$

$PE\,(U)$	Partitionsentropie zur Beurteilung der Clustervalidität
q	Parameter zur Kompensierung von Parallelität und Distanz bei Bestimmung der Clusterähnlichkeit
q_1	Index für Attribute in der Bedingung einer Assoziationsregel
q_2	Index für Attribute in der Folgerung einer Assoziationsregel
Q	Menge der kompatiblen Cluster bei der Clustervereinigung
r	relevante Assoziationsregel
r_i	i-te Assoziationsregel, $i = 1, \ldots, c$
R	Regelbasis, d.h. Menge der relevanten Regeln; $R = \{r_1, \ldots, r_c\}$
RMT	Rule Matching Threshold (RMT), d.h. Grenzwert zur Überprüfung der Übereinstimmung von Assoziationsregeln
$s_{ii'}$	Ähnlichkeit zwischen zwei Clustern C_i und $C_{i'}$ nach Aggregation ihrer Distanz- und Parallelitätswerte
$s_{ii'}^{AR}$	Ähnlichkeit zwischen zwei Assoziationsregel r_i und $r_{i'}$
$s_{i\cdot}^{AR}$	maximale Ähnlichkeit einer Assoziationsregel r_i aus Periode t verglichen mit allen Assoziationsregeln aus Periode $t^{'}$
$s_{\cdot i'}^{AR}$	maximale Ähnlichkeit einer Assoziationsregel $r_{i'}$ aus Periode $t^{'}$ vergleichen mit allen Assoziationsregeln aus Periode t
$s_{ii'}^{\mathtt{I}}$	Ähnlichkeit zwischen zwei Clustern C_i und $C_{i'}$ auf Basis des Inklusionsmaßes $\mathtt{I}_{ii'}$
S_i	Annäherung des Vielfachen der Kovarianzmatrix Σ_i von Cluster C_i
$S\,(U)$	Separationsindex zur Beurteilung der Clustervalidität
$St\,(c_t)$	Strukturstärke einer Clusterstruktur bei c_t Clustern, d.h. Maß zur Bestimmung der Effektivität der Clusterzahl und ihrer Genauigkeit
t	Index der Perioden, $t = 1, \ldots, T$
t_i^0	Zeitpunkt des erstmaligen Auftretens des Ausreißerclusters $C_{t_i}^{\mathtt{A}}$
$\hat{t}_i^{CT}$	Zeitpunkt für die erwartete Teilung von Cluster C_i
$\hat{t}_i^{CT'}$	Zeitpunkt für eine als ausreichend erwartete Separation der Subcluster auf Basis der Distanz der Subclusterzentren
$t_{eC}^{\max}$	Zeitdauer, über die Grenzwert $\lambda_{\min}^{neu}$ maximal unterschritten werden darf, bevor ein Cluster eliminiert wird

$t_{nC}^{\max}$	Zeitdauer, über die Grenzwert $\delta_{\min}$ bei Entstehung eines neuen Clusters maximal unterschritten werden darf
$\hat{t}_{ii'}^{V}$	Prognostizierter Zeitpunkt für Vereinigung der Cluster C_i und $C_{i'}$
Δt	Parameter für die Länge des Analyseintervalls, d.h. nach wie vielen Perioden eine erneute Analyse im dynamischen Kontext durchgeführt werden soll
T	Periodenzahl
$\mathfrak{T}$	Menge aller Transaktionen (Association Rule Mining)
$\mathfrak{T}_A$	Menge aller Transaktionen, die Itemmenge A enthalten
$\mathfrak{T}_t$	Menge aller Transaktionen zum Zeitpunkt t
u	Fuzzy-Menge; Funktion von einer Grundmenge O in das Einheitsintervall
u_{ij}	Zugehörigkeitsgrad von Objekt $\vec{x}_j$ zu Cluster C_i
$u_{ii'}^{1}$	Zugehörigkeitsgrad der Cluster C_i und $C_{i'}$ bzgl. ihrer Distanz zur Fuzzy-Menge *nahe 0*
$u_{ii'}^{2}$	Zugehörigkeitsgrad der Cluster C_i und $C_{i'}$ bzgl. ihrer Parallelität zur Fuzzy-Menge *nahe 1*
U	$c \times n$-Matrix der Zugehörigkeitsgrade u_{ij}
$[U]_\alpha$	α-Schnitt von u, d.h. Menge aller Objekte mit einem Zugehörigkeitsgrad von mindestens α
$\vec{v}_i$	Eigenschaftsvektor des Zentrums von Cluster C_i
$\vec{v}_i^{neu}$	ungewichtetes Zentrum der von Cluster C_i neu absorbierten Objekte
$\hat{\vec{v}}_i$	Schätzung des aktuellen Clusterzentrums von Cluster C_i
Δv_{il}	Veränderung des Clusterzentrums von Cluster C_i bzgl. Dimension l im Vergleich zur vorherigen Analyse
w	Anzahl Items (Association Rule Mining)
$\vec{x}_j$	Eigenschaftsvektor eines Objektes j
x_{jl}	Eigenschaftsausprägung des j-ten Objekts, $j = 1, \ldots, n$, in der l-ten Dimension, $l = 1, \ldots, p$
$\vec{x}_{t_j}^{\mathrm{A}}$	Eigenschaftsvektor einer Ausreißerobjektes j zum Zeitpunkt t
X	$n \times p$-Matrix der Objekte bei n Objekten und p Dimensionen
$y_{ii'q_1}$	Binärvariable zur Bestimmung der Übereinstimmung des q_1-ten Attributs der Bedingungen zweier Assoziationsregeln r_i und $r_{i'}$

Y_i	Teilcluster, das zur Berechnung der mittleren Partitionsdichte $APD\,(U,C)$ benötigt wird; $Y_i = \left\{\vec{x}_j \middle\vert\, (\vec{x}_j - \vec{v}_i)^{\mathrm{T}}\, \Sigma_i^{-1}\, (\vec{x}_j - \vec{v}_i) < 1, j \in \{1,\ldots,n\}\right\}$
$z_{ii'q_2}$	Binärvariable zur Bestimmung der Übereinstimmung des q_2-ten Attributs der Folgerungen zweiter Assoziationsregeln r_i und $r_{i'}$
α	Grenzwert der Absorbierung zur Bestimmung des α-Schnitts im Bestrafungsterm
α^{A}	Grenzwert der Absorbierung für Analyse struktureller Änderungen innerhalb der Clusterstruktur
$\alpha_\kappa^{\mathrm{A}}$	Grenzwert zur Differenzierung zwischen möglicher Clusterelimination und -trennung auf Basis des Kompaktheitsindexes $\kappa_i^{\alpha^{\mathrm{A}}}$; $\alpha_\kappa^{\mathrm{A}} \in (\alpha^{\mathrm{A}}, 1]$
α_{SC}^{A}	Grenzwert der Absorption für Extraktion der zum Subclustering verwendeten Objekte; $\alpha_{SC}^{\mathrm{A}} \in [0, \alpha^{\mathrm{A}})$
α^{AR}	Grenzwert für Menge der beizubehaltenen Vergangenheitsinformationen (Association Rule Mining)
$\beta^{\vec{v}}$	Parameter zur Bestimmung des geforderten Anteils der Fuzzy-Varianz bei einer Clusterverschiebung
$\beta_{CT}^{\min}$	Parameter zur Bestimmung der geforderten Mindestgröße und -dichte eines Subclusters für eine potentielle Clustertrennung; $\beta_{CT}^{\min} \in [0,1]$
β_{CT}	Parameter zur Bestimmung der geforderten Mindestgröße und -dichte eines Subclusters für eine vollzogene Clustertrennung; $\beta_{CT} \in \left[\beta_{CT}^{\min}, 1\right]$
$\beta_{eC}^{\min}$	Parameter zur Bestimmung der geforderten Mindestgröße und -dichte eines noch existierenden Clusters; $\beta_{eC}^{\min} \in [0,1]$
β_{eC}	Parameter zur Bestimmung der geforderten Mindestgröße und -dichte eines ungefährdeten Clusters; $\beta_{eC} \in \left[\beta_{eC}^{\min}, 1\right]$
$\beta_{nC}^{\min}$	Parameter zur Bestimmung der geforderten Mindestgröße und -dichte eines potentiell entstehenden Clusters; $\beta_{nC}^{\min} \in [0,1]$
β_{nC}	Parameter zur Bestimmung der geforderten Mindestgröße und -dichte eines bereits entstandenen Clusters; $\beta_{nC} \in \left[\beta_{nC}^{\min}, 1\right]$
γ	allgemeiner Gewichtungsparameter
γ_i	clusterspezifischer Gewichtungsparameter für Cluster C_i
γ^{Σ}	Parameter für Gewichtung neuer Objekte bei Berechnung der γ_i^{Σ}
γ_i^{Σ}	Gewicht für Einfluss neuer Objekte bei Update der Kovarianzmatrix

$\gamma^{\vec{v}}$	Parameter für Gewichtung neuer Objekte bei Berechnung der $\gamma_i^{\vec{v}}$
$\gamma_i^{\vec{v}}$	Gewicht für Einfluss neuer Objekte bei einer Clusterverschiebung
γ_{poss}	Gewichtungsparameter für possibilistische Zugehörigkeitsgrade bei kombinierter Optimierungsfunktion
γ_{prob}	Gewichtungsparameter für probabilistische Zugehörigkeitsgrade bei kombinierter Optimierungsfunktion
γ^s	Gewichtung der Distanz im Vergleich zur Parallelität bei Aggregation der Werte zur Bestimmung der Ähnlichkeit zweier Cluster
γ^{St}	Gewichtungsparameter zur Gewichtung der Ziele bei der Berechnung der Strukturstärke $St(c_t)$
$\delta_{\min}$	Grenzwert für das minimale Wachstum eines Ausreißerclusters, um weiterhin als potentielles Cluster angesehen zu werden
$\delta(\vec{x}_j, \vec{x}_k)$	Monotonieanpassung bzgl. der Objektvektoren $\vec{x}_j$ und $\vec{x}_k$ bei der nichtmetrischen MDS
$\Delta\left(r_i^t, r_{i'}^{t'}\right)$	Differenzmaß für zwei Assoziationsregeln r_i^t und $r_{i'}^{t'}$
$\Delta'\left(r_i^t, r_{i'}^{t'}\right)$	modifiziertes Differenzmaß für zwei Assoziationsregeln r_i^t und $r_{i'}^{t'}$
ϵ	Abbruchkriterium bei iterativen Bestimmungen innerhalb von Algorithmen
ζ	allgemeiner Normierungsparameter
ζ_i	clusterspezifischer Normierungsparameter für Bestrafungsterm bei Ansätzen zur Modellierung der Clusterhomogenität
η_i	Schätzer für die Ausdehnung von Cluster C_i
θ_{il}	l-ter Eigenwert der Kovarianzmatrix von Cluster C_i, $l = 1, \cdots, p$
ι	Item (Assoziation Rule Mining)
$\kappa_i^{\alpha^{\mathrm{A}}}$	Kompaktheitsindex zur Bestimmung der (normierten) Fuzzy-Varianz innerhalb des Clusters
λ_{dist}^{CT}	Grenzwert für die minimal benötigte Distanz zwischen Subclusterzentren, um als geteilt gelten zu können
λ_{par}^{CT}	Grenzwert für minimal benötigte Clusterdrehung für eine mögliche Clustertrennung; $\lambda_{par}^{CT} \in [0, 1]$
λ_{θ}^{CT}	Grenzwert für minimal geforderte Ausrichtungsänderung für eine mögliche Clustertrennung; $\lambda_{\theta}^{CT} > 0$
$\lambda_{CT}^{\mathrm{I}^{\max}}$	Grenzwert für die maximal zulässige Überschneidung zweier Subcluster, um als potentiell teilbar zu gelten; $\lambda_{CT}^{\mathrm{I}^{\max}} \in [0, 1]$
$\lambda_{CT}^{\mathrm{I}^{\min}}$	Grenzwert für die maximal zulässige Überschneidung zweier Subcluster, um als bereits geteilt zu gelten; $\lambda_{CT}^{\mathrm{I}^{\min}} \in \left[0, \lambda_{CT}^{\mathrm{I}^{\max}}\right]$

$\lambda_V^{\mathtt{I}^{\max}}$	Grenzwert für Überschneidungsgrad, ab dem eine Vereinigung zweier Cluster ohne weitere Prüfung der Distanz- und Parallelitätswerte erfolgt; $\lambda_{\max}^{\mathtt{I}} \in [\lambda_{\min}^{\mathtt{I}}, 1]$
$\lambda_V^{\mathtt{I}^{\min}}$	Grenzwert für minimalen Überschneidungsgrad, ab dem für eine Vereinigung zweier Cluster eine weitere Prüfung der Distanz- und Parallelitätswerte erfolgt; $\lambda_{\min}^{\mathtt{I}} \in [0, 1]$
$\lambda_{\min}^{neu}$	Grenzwert für die minimale Neuzuordnung von Objekten je Periode
$\lambda_{\max}^{s}$	Grenzwert für Ähnlichkeit zwischen zwei Clustern nach Aggregation ihrer Distanz- und Parallelitätswerte, ab dem eine Vereinigung zweier Cluster erfolgt; $\lambda_{\max}^{s} \in [\lambda_{\min}^{s}, 1]$
$\lambda_{\min}^{s}$	Grenzwert für minimale Ähnlichkeit zwischen zwei Clustern nach Aggregation ihrer Distanz- und Parallelitätswerte, ab dem eine zukünftige Vereinigung als möglich gilt; $\lambda_{\min}^{s} \in [0, 1]$
λ_{dist}^{V}	Grenzwert für maximale Distanz bei Clustervereinigung auf Basis der Distanz zweier Zentren
λ_{par}^{V}	Grenzwert für minimale Parallelität bei Clustervereinigung auf Basis der Parallelität zweier Clusterprototypen
μ	vernachlässigbar kleiner Wert zur Verhinderung der Division mit Null
ν_1	Support für die Fuzzy-Menge *nahe 0*
ν_2	Support für die Fuzzy-Menge *nahe 1*
$\vec{\xi}$	Parameter, verwendet zur Herleitung der Updateregeln
π	Parameter bei Bestimmung abnehmend- bzw. zunehmend-häufiger Itemmengen zur Festlegung, ob A häufige Itemmenge ist
ρ	Iterationsindex für Algorithmen
σ_{ill}^2	Fuzzy-Varianz von $[U_i]_\alpha$ der l-ten Dimension
Σ_i	Kovarianzmatrix von Cluster C_i
Σ_i^{hart}	harte Kovarianzmatrix des α^{A}-Schnitts von Cluster C_i
$\hat{\Sigma}_i^{hart}$	Schätzung der aktuellen harten Kovarianzmatrix des α^{A}-Schnitts von Cluster C_i
Σ_i^{neu}	harte Kovarianzmatrix des α^{A}-Schnitts der von Cluster C_i neu absorbierten Objekte
τ	Parameter für die Länge des betrachteten Zeitintervalls
υ_t	Index der Perioden für Teilintervall, $\upsilon_1 = t_1, \ldots, t_2$
ϕ	Gewichtungsfaktor für den Einbeziehungsgrad der Anpassung bei simultaner Dynamisierung (MDS)
$\chi_{ii'}$	Maß zur Bestimmung der Signifikanz bei unerwarteten Änderungen von Assoziationsregeln
ψ_j	j-te Transaktion in $\mathfrak{T}$

$\omega_{ii'}$	Binärvariable zur Unterscheidung zwischen *emerging patterns* und unerwarteten Regel (Association Rule Mining)

Literaturverzeichnis

[Aggarwal u. a. 2003] Aggarwal, C.C. ; Han, J. ; Wang, J. ; Yu, P.S.: A Framework for Clustering Evolving Data Streams. In: *Proceedings of the 29th International Conference on Very Large Databases (VLDB'03)*. Berlin, 2003, S. 81–92

[Agrawal u. a. 1993] Agrawal, R. ; Imilienski, T. ; Swami, A.: Mining Association Rules between Sets of Items in Large Databases. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data*. Washington, D.C., 1993, S. 207–216

[Agrawal und Psaila 1995] Agrawal, R. ; Psaila, G.: Active Data Mining. In: *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD'95)*. Montreal, 1995, S. 3–8

[Agrawal und Srikant 1994] Agrawal, R. ; Srikant, R.: Fast Algorithms for Mining Association Rules. In: *Proceedings of the 20th VLDB Conference*. Santiago de Chile, 1994, S. 487–499

[Ali und Ketchpel 2003] Ali, K. ; Ketchpel, S.P.: Golden Path Analyzer: Using Divide-and-Conquer to Cluster Webs. In: *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Washington,D.C., 2003, S. 349–358

[Ambrosi und Hansohm 1986] Ambrosi, K. ; Hansohm, J.: Ein dynamischer Ansatz zur Repräsentation von Objekten. In: *Operations Research Proceedings – Vorträge der 15. Jahrestagung 1986*. Ulm, 1986, S. 425–431

[Angers 2002] Angers, J.-F.: Curves Comparison Using Wavelets / Université de Montréal, Départment de mathématiques et de statistique. 2002. – Forschungsbericht

[Angstenberger 2000] Angstenberger, L.: *Dynamic Fuzzy Pattern Recognition*, RWTH Aachen, Dissertation, 2000

[Antunes und Oliveira 2001] Antunes, C.M. ; Oliveira, A.L.: Temporal Data Mining: An Overview. In: *Workshop on Temporal Data Mining at the 7th International Conference on Knowledge Discovery and Data Mining (KDD'01)*. San Francisco, CA, 2001

[Apeh und Gabrys 2011] Apeh, E. ; Gabrys, B.: Change Mining of Customer Profiles Based on Transactional Data. In: *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on Data Mining*. Vancouver, 2011, S. 560–567

[Asuncion und Newman 2007] Asuncion, A. ; Newman, D.J.: *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Science, 2007. – `http://www.ics.uci.edu/~mlearn/MLRepository.html`

[AT&T 2013] AT&T: *Corporate Profile*, 2013. – `http://www.att.com/gen/investor-relations?pid=5711`, Verifizierungsdatum: 24.04.2013

[Bacher u. a. 2010] Bacher, J. ; Pöge, A. ; Venzig, K.: *Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren*. München : Oldenbourg Verlag, 2010

[Backhaus u. a. 2006] Backhaus, K. ; Erichson, B. ; Plinke, W. ; Weiber, R.: *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung.* Berlin u.a. : Springer, 2006. – 11. Auflage

[Baier und Brusch 2008] Baier, D. ; Brusch, M.: Marktsegmentierung. In: Herrmann, A. (Hrsg.) ; Homburg, C. (Hrsg.) ; Klarmann, M. (Hrsg.): *Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele.* Wiesbaden : Gabler, 2008, S. 769–790. – 3. Auflage

[Banerjee und Ghosh 2001] Banerjee, A. ; Ghosh, J.: Clickstream Clustering Using Weighted Longest Common Subsequences. In: *Proceedings of the Web Mining Workshop at the 1st SIAM Conference on Data Mining*, 2001, S. 33–40

[Bay und Pazzani 1999] Bay, S.D. ; Pazzani, M.J.: Detecting Change in Categorical Data: Mining Contrast Sets. In: *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining (KDD-99).* San Diego, California, 1999, S. 302–306

[Bensaid u. a. 1996] Bensaid, A.M. ; Hall, L.O. ; Bezdek, J.C. ; Clarke, L.P. ; Silbiger, M.L. ; Arrington, J.A. ; Murtagh, R.F.: Validity-Guided (Re)Clustering with Applications to Image Segmentation. In: *IEEE Transactions on Fuzzy Systems* 4/2 (1996), S. 112–123

[Berekoven u. a. 2006] Berekoven, L. ; Eckert, W. ; Ellenrieder, P.: *Marktforschung – Methodische Grundlagen und praktische Anwendung.* Wiesbaden : Gabler, 2006. – 11. Auflage

[Berry und Linoff 2000] Berry, M.J.A. ; Linoff, G.S.: *Mastering Data Mining – The Art and Science of Customer Relationship Management.* New York : Wiley, 2000

[Bezdek 1980] Bezdek, J.C.: Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2/1 (1980), S. 1–8

[Bezdek 1981] Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms.* New York : Plenum, 1981

[Borg und Groenen 2005] Borg, I. ; Groenen, P.J.E.: *Modern Multidimensional Scaling – Theory and Applications.* New York : Springer, 2005. – 2. Auflage

[Borgelt 2005] Borgelt, C.: *Prototype-based Classification and Clustering.* Habilitation. 2005

[Böttcher u. a. 2008] Böttcher, M. ; Höppner, F. ; Spiliopoulou, M.: On Exploiting the Power of Time in Data Mining. In: *ACM SIGKDD Explorations Newsletter* 10/2 (2008), S. 3–11

[Böttcher u. a. 2009] Böttcher, M. ; Spott, M. ; Nauck, D. ; Kruse, R.: Mining Changing Customer Segments in Dynamic Markets. In: *Expert Systems with Applications* 36/1 (2009), S. 155–164

[Bramer 2007] Bramer, M.: *Principles of Data Mining.* London : Springer, 2007

[Cao u. a. 2006] Cao, F. ; Ester, M. ; Qian, W. ; Zhou, A.: Density-Based Clustering over an Evolving Data Stream with Noise. In: *Proceedings of the 6th SIAM International Conference on Data Mining*. Bethesda, MD, 2006, S. 326–337

[Carroll und Chang 1970] Carroll, J.D. ; Chang, J.J: Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. In: *Psychometrika* 35 (1970), S. 283–320

[Chakrabarti u. a. 1998] Chakrabarti, S. ; Sarawagi, S. ; Dom, B.: Mining Suprising Patterns Using Temporal Description Length. In: *Proceedings of the 24th VLDB Conference*. New York, 1998, S. 606–617

[Chen u. a. 2005] Chen, M.-C. ; Chiu, A.-L. ; Chang, H.-H.: Mining Changes in Customer Behavior in Retail Marketing. In: *Expert Systems with Applications* 28 (2005), S. 773–781

[Cios u. a. 2007] Cios, K.J. ; Pedrycz, W. ; Swiniarski, R.W. ; Kurgan, L.A.: *Data Mining – A Knowledge Discovery Approach*. New York : Springer, 2007

[Crespo und Weber 2005] Crespo, F. ; Weber, R.: A Methodology for Dynamic Data Mining Based on Fuzzy Clustering. In: *Fuzzy Sets and Systems* 150 (2005), S. 267–284

[Decker 1993] Decker, R.: *Analyse und Simulation des Kaufverhaltens auf Konsumgütermärkten*. Frankfurt am Main : Peter Lang – Europäischer Verlag der Wissenschaften, 1993

[Decker 1994] Decker, R.: Ein wissensbasiertes System zur Kaufverhaltensanalyse – WIMDAS-KVA. In: Gaul, W. (Hrsg.) ; Schader, M. (Hrsg.): *Wissensbasierte Marketing-Datenanalyse*. Frankfurt am Main : Peter Lang – Europäischer Verlag der Wissenschaften, 1994, Kap. 6, S. 113–144

[Deichsel und Trampisch 1985] Deichsel, G. ; Trampisch, H.J.: *Clusteranalyse und Diskriminanzanalyse*. Stuttgart, New York : Gustav Fischer Verlag, 1985

[Deimer 1986] Deimer, R.: *Unscharfe Clusteranalysemethoden*. Idstein : Schulz-Kirchner-Verlag, 1986

[Dong u. a. 2003] Dong, G. ; Han, J. ; Lakshmanan, L.V.S. ; Pei, J. ; Wong, H. ; Yu, P.S.: Online Mining of Changes from Data Streams: Research Problems and Preliminary Results. In: *Proceedings of the 2003 ACM SIGMOD Workshop on Management and Processing of Data Streams (MPDS'03)*. San Diego, California, 2003

[Dong und Li 1999] Dong, G. ; Li, J: Efficient Mining of Emerging Patterns: Discovering Trends and Differences. In: *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining (KDD-99)*. San Diego, CA, 1999, S. 43–52

[Dunn 1973] Dunn, J.C.: A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact, Well Separated Clusters. In: *Journal of Cybernetics* 3/3 (1973), S. 32–57

[Fayyad u. a. 1996] Fayyad, U. ; Piatetsky-Shapiro, G. ; Smyth, P.: From Data Mining to Knowledge Discovery in Databases. In: *AI Magazin* Fall 1996 (1996), S. 37–54

[Freitas 2002] Freitas, A.A.: A survey of evolutionary algorithms for data mining and knowledge discovery. In: Ghosh, A. (Hrsg.) ; Tsutsui, S. (Hrsg.): *Advances in evolutionary computing: theory and applications*. New York, NY : Springer, 2002, Kap. 33, S. 819–845

[Frigui und Krishnapuram 1997] Frigui, H. ; Krishnapuram, R.: Clustering by Competitive Agglomeration. In: *Pattern Recognition* 30/7 (1997), S. 1109–1119

[Gath und Geva 1989] Gath, I. ; Geva, A.B.: Unsupervised Optimal Fuzzy Clustering. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (1989), S. 773–781

[Gower und Dijksterhuis 2004] Gower, J.C. ; Dijksterhuis, G.B.: *Procrustes Problems*. Oxford : Oxford University Press, 2004

[Green u. a. 1989] Green, P.E. ; Carmone, F.J. ; Smith, S.M.: *Multidimensional Scaling – Concepts and Applications*. Boston u.a. : Allyn and Bacon, 1989

[Gudmundsson u. a. 2008] Gudmundsson, J. ; Laube, P. ; Wolle, T.: Movement Patterns in Spatio-Temporal Data. In: Shekhar, S. (Hrsg.) ; Xiong, H. (Hrsg.): *Encyclopedia of GIS*. Heidelberg : Springer, 2008, S. 726–732

[Gustafson und Kessel 1979] Gustafson, E.E. ; Kessel, W.C.: Fuzzy Clustering with a Fuzzy Covariance Matrix. In: *Proceedings IEEE CDC*. San Diego, 1979, S. 761–766

[Höppner u. a. 1999] Höppner, F. ; Klawonn, F. ; Kruse, R. ; Runkler, T.: *Fuzzy Cluster Analysis – Methods for Classification, Data Analysis and Image Recognition*. Chichester : Wiley, 1999

[Jensen 2008] Jensen, O.: Clusteranalyse. In: Herrmann, A. (Hrsg.) ; Homburg, C. (Hrsg.) ; Klarmann, M. (Hrsg.): *Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele*. Wiesbaden : Gabler, 2008, S. 335–372. – 3. Auflage

[Joentgen u. a. 1999] Joentgen, A. ; Mikenina, L. ; Weber, R. ; Zimmermann, H.-J.: Dynamic Fuzzy Data Analysis Based on Similarity Between Functions. In: *Fuzzy Sets and Systems* 105 (1999), S. 81–90

[Kalnis u. a. 2005] Kalnis, P. ; Mamoulis, N. ; Bakiras, S.: On Discovering Moving Clusters in Spatio-temporal Data. In: *Proceedings of the 9th International Symposium on Advances in Spatial and Temporal Databases (SSTD'05)*. Angra dos Reis, 2005, S. 364–381

[Kaymak 1998] Kaymak, U.: *Fuzzy Decision Making with Control Applications*, Delft University of Technology, Dissertation, 1998

[Kaymak und Babuška 1995] Kaymak, U. ; Babuška, R.: Compatible Cluster Merging for Fuzzy Modelling. In: *Proceedings of the Fourth IEEE International Conference on Fuzzy Systems* Bd. 2, 1995, S. 897–904

[Klawonn und Höppner 2003] Klawonn, F. ; Höppner, F.: What is Fuzzy About Fuzzy Clustering? Understanding and Improving the Concept of the Fuzzifier. In: *Cryptographic Hardware and Embedded Systems – CHES 2003*. Berlin u.a. : Springer, 2003, S. 254–264. – Lecture Notes in Computer Science

[Krishnapuram und Freg 1992] Krishnapuram, R. ; Freg, C.P.: Fitting an Unknown Number of Lines and Planes to Image Data Through Compatible Cluster Merging. In: *Pattern Recognition* 25/4 (1992), S. 385–400

[Krishnapuram und Keller 1993] Krishnapuram, R. ; Keller, J.M.: A Possibilistic Approach to Clustering. In: *IEEE Transactions on Fuzzy Systems* 1 (1993), S. 98–110

[Kruse u. a. 2007] Kruse, R. ; Döring, C. ; Lesot, M.-J.: Fundamentals of Fuzzy Clustering. In: Oliveira, J. Valente de (Hrsg.) ; Pedrycz, W. (Hrsg.): *Advances in Fuzzy Clustering and its Applications*. Chichester : Wiley, 2007, Kap. 1, S. 3–30

[Kruse u. a. 1993] Kruse, R. ; Gebhardt, J. ; Klawonn, F.: *Fuzzy-Systeme*. Stuttgart : B.G. Teubner, 1993

[Kruskal 1964] Kruskal, J.B.: Nonmetric Multidimensional Scaling: A Numerical Method. In: *Psychometrika* 29/2 (1964), S. 115–159

[Lesot und Kruse 2006] Lesot, M.-J. ; Kruse, R.: Gustafson-Kessel-like Clustering Algorithm Based on Typicality Degrees. In: *Proceedings of IPMU'06*. Paris, 2006, S. 1300–1307

[Li u. a. 2012] Li, I.-H. ; Huang, J.-Y. ; Liao, I.-E.: Predicting Sequential Pattern Changes in Data Streams. In: *International Journal of Innovative Computing, Information and Control* 8/1 (2012), S. 286–302

[Li und Mukaidono 1995] Li, R.-P. ; Mukaidono, M.: A Maximum-Entropy Approach to Fuzzy Clustering. In: *Proceedings of the International Joint Conference, 4th IEEE International Conference Fuzzy/2nd International Fuzzy Engineering Symposium (FUZZ/IEEE-IFES)*. Yokohama, 1995, S. 2227–2232

[Liu und Hsu 1996] Liu, B. ; Hsu, W.: Post-analysis of Learned Rules. In: *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI-96)*. Portland, OR, 1996, S. 828–834

[Mahalanobis 1930] Mahalanobis, P.C.: On the Tests and Measures of Group Divergences. In: *Journal of the Asiatic Society of Benegal* 26/4 (1930), S. 541–588

[Mahalanobis 1936] Mahalanobis, P.C.: On the Generalised Distance in Statistics. In: *Proceedings of the National Institute of Sciences of India*, 1936, S. 49–55

[Mathar 1997] Mathar, R.: *Multidimensionale Skalierung*. Stuttgart : B.G. Teuber, 1997

[Minke und Ambrosi 2012] Minke, A. ; Ambrosi, K.: Predicting Changes in Market Segments Based on Customer Behavior. In: *Proceedings of the 36th Annual Conference of the German Classification Society (GfKl'12)*. Hildesheim, 2012. – zur Veröffentlichung angenommen

[Minke u. a. 2009] Minke, A. ; Ambrosi, K. ; Hahne, F.: Approach for Dynamic Problems in Clustering. In: *Proceedings of the 4th International Symposium on Information Technologies in Environmental Engineering (ITEE'09)*. Thessaloniki, 2009, S. 373–386

[Minke und Lessing 2010] Minke, A. ; Lessing, H.: Environmental Monitoring, Data Mining, and Dynamic Analysis. In: Teuteberg, F. (Hrsg.) ; Marx Gomez, J. (Hrsg.): *Corporate Environmental Management Information Systems: Advancements and Trends*. Hershey, PA : IGI Global, 2010, Kap. 11, S. 168–179

[Nasraoui u. a. 2003] Nasraoui, O. ; Uribe, C.C. ; Coronel, C.R. ; Gonzalez, F.: TECNO-STREAMS: Tracking Evolving Clusters in Noisy Data Streams with a Scalable Immune System Learning Model. In: *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03)*. Melbourne, 2003, S. 235–242

[Nassar u. a. 2004] Nassar, S. ; Sander, J. ; Cheng, C.: Incremental and Effective Data Summarization for Dynamic Hierarchical Clustering. In: *Proceedings of the ACM SIGMOD Conference*. Paris, 2004, S. 467–478

[Neumann 2006] Neumann, A.: *Individuelle und dynamische multidimensionale Skalierung – mit Anwendung in der Marktforschung*, Universität Hildesheim, Masterarbeit, 2006

[Neumann 2008] Neumann, A.: Fuzzy Clustering in Dynamic Problems / Universität Hildesheim, Institut für Betriebswirtschaft und Wirtschaftsinformatik. 2008. – Forschungsbericht

[Valente de Oliveira und Pedrycz 2007] Oliveira, J. Valente de (Hrsg.) ; Pedrycz, W. (Hrsg.): *Advances in Fuzzy Clustering and its Applications*. Chichester : Wiley, 2007

[Pal u. a. 1997] Pal, N. ; Pal, K. ; Bezdek, J.: A Mixed C-Means Clustering Model. In: *Fuzz-IEEE'97*, 1997, S. 11–21

[Pal u. a. 2004] Pal, N. ; Pal, K. ; Keller, J.M. ; Bezdek, J.: A New Hybrid C-Means Clustering Model. In: *Fuzz-IEEE'04*, 2004, S. 179–184

[Pechoucek u. a. 1999] Pechoucek, M. ; Štepánková, O. ; Mikšovský, P.: Maintenance of Discovered Knowledge. In: *Proceedings of the 3rd European Conference on Principles of Data Mining and Knowledge Discovery*. Prague, 1999, S. 476–483. – Lecture Notes in Computer Science

[Raghavan und Hafez 2000] Raghavan, V. ; Hafez, A.: Dynamic Data Mining. In: *Proceedings of the 13th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*. New Orleans, LA, 2000, S. 220–229

[Rissland und Friedman 1995] Rissland, E.L. ; Friedman, M.T.: Detecting Change in Legal Concepts. In: *Proceedings of the 5th International Conference on Artificial Intelligence and Law*. College Park, 1995, S. 127–136

[Sanz Sáiz 2005] Sanz Sáiz, B.: Data mining for Advanced Customer Management. In: *Proceedings ASMDA 2005*. Brest, 2005, S. 325–333

[Schobert 1979] Schobert, R.: *Die Dynamisierung komplexer Marktmodelle mit Hilfe von Verfahren der Mehrdimensionalen Skalierung*. Berlin : Duncker & Humboldt, 1979

[Schobert und Dichtl 1979] Schobert, R. ; Dichtl, E.: *Mehrdimensionale Skalierung*. München : Verlag Vahlen, 1979

[Setnes und Kaymak 1998] Setnes, M. ; Kaymak, U.: Extended Fuzzy C-Means with Volume Prototypes and Cluster Merging. In: *Proceedings EUFIT'98*. Aachen, 1998, S. 1360–1364

[Shepard 1957] Shepard, R.: Stimulus and Response Generalization – A Stochastic Model Relating Generalization to Distance in Psychological Space. In: *Psychometrika* 22/4 (1957), S. 325–345

[Silberschatz und Tuzhilin 1996] Silberschatz, A. ; Tuzhilin, A.: What Makes Patterns Interesting in Knowledge Discovery Systems? In: *IEEE Transactions on Knowledge and Data Engineering* 8 (6) (1996), S. 970–974

[Song u. a. 2001] Song, H.S. ; Kim, J.K. ; Kim, S.H.: Mining the Change of Customer Behavior in an Internet Shopping Mall. In: *Expert Systems with Applications* 21 (2001), S. 157–168

[Stutz 1998] Stutz, C.: Partially Supervised Fuzzy C-Means Clustering with Cluster Merging. In: *Proceedings EUFIT'98*. Aachen, 1998, S. 1725–1729

[Tan u. a. 2013] Tan, T. ; Suk, H.W. ; Hwang, H. ; Lim, J.: Functional Fuzzy Clusterwise Regression Analysis. In: *Advances in Data Analysis and Classification* 7 (2013), S. 57–82

[Timm 2002] Timm, H.: *Fuzzy-Clusteranalyse: Methoden zur Exploration von Daten mit fehlenden Werten sowie klassifizierten Daten*, Otto-von-Guericke-Universität Magdeburg, Dissertation, 2002

[Timm u. a. 2001] Timm, H. ; Borgelt, C. ; Döring, C. ; Kruse, R.: Fuzzy Cluster Analysis with Cluster Repulsion. In: *Proceedings of the European Symposium on Intelligent Technologies (EUNITE)*. Tenerife, 2001. – CD-ROM

[Wang u. a. 2003] Wang, K. ; Zhou, S. ; Fu, C.A. ; Yu, J.X.: Mining Changes of Classification by Correspondence Tracing. In: *Proceedings of the 3rd SIAM International Conference on Data Mining* Bd. 3. San Francisco, California, 2003, S. 95–106

[Weber 2007] Weber, R.: Fuzzy Clustering in Dynamic Data Mining – Techniques and Applications. In: Oliviera, J. Valente de (Hrsg.) ; Pedrycz, W. (Hrsg.): *Advances in Fuzzy Clustering and its Applications*. Chichester : Wiley, 2007, Kap. 15, S. 315–332

[Wiedmann u. a. 2003] Wiedmann, K.-P. ; Buckler, F. ; Buxel, H.: Data Mining – ein einführender Überblick. In: Wiedmann, K.-P. (Hrsg.) ; Buckler, F. (Hrsg.): *Neuronale Netze im Markting-Management – Praxisorientierte Einführung in modernes Data-Mining*. Wiesbaden : Gabler, 2003, Kap. 1, S. 19–37. – 2. Auflage

[Xie und Beni 1991] Xie, X.L. ; Beni, G.: A Validity Measure for Fuzzy Clustering. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (1991), S. 841–874

[Zadeh 1965] Zadeh, L.A.: Fuzzy Sets. In: *Information and Control* 8 (1965), S. 338–353

[Zadeh 1978] Zadeh, L.A.: Fuzzy Sets as a Basis for a Theory of Possibility. In: *Fuzzy Sets and Systems* 1 (1978), S. 3–28

[Zhou u. a. 2008] Zhou, A. ; Cao, F. ; Qian, W. ; Jin, C.: Tracking Clusters in Evolving Data Streams over Sliding Windows. In: *Knowledge and Information Systems* 15 (2008), S. 181–214